On a non-parametric confidence interval for the regression slope
(Running title: A la Tukey confidence interval for slope)

Róbert Tóth and Ján Somorčík*
Department of Applied Mathematics and Statistics
Faculty of Mathematics, Physics and Informatics
Comenius University Bratislava
Mlynská dolina, 842 48 Bratislava, Slovakia

Abstract
We investigate the application of the Tukey’s methodology in Theil’s regression to obtain a confidence interval for the true slope in the straight line regression model. We illustrate by Monte Carlo simulations, that this approach, unlike the classical Theil’s approach based on Kendall’s tau, deflates the true confidence level of the resulting interval. We provide also rigorous proofs in case of $n = 4$ data points (in general) and in case of $n = 5$ data points (under uniform distribution of errors).

Keywords: Theil’s regression; Tukey’s confidence interval; Walsh averages; software R.

1 Introduction
The Theil’s regression is a robust non-parametric replacement of the traditional least squares approach to the straight line regression model $Y = \beta_0 + \beta_1 x + \varepsilon$ and also to some more complex linear regression models (the pioneering papers were Theil, 1950a, Theil, 1950b, and Theil, 1950c). The Theil’s methodology does not require normality of the random errors $\varepsilon$, while being able to provide parameter estimates, tests of linear hypotheses about the parameters, as well as confidence intervals for the parameters (see e.g. Hollander & Wolfe, 1999 for a detailed description of the methods).

Our paper focuses on the confidence interval for the true slope $\beta_1$. The classical Theil’s confidence interval proposed already in Theil (1950a) makes use of the theory of Kendall’s tau. However, recently, we came across an approach that utilizes the well-known confidence interval based on the Wilcoxon signed rank test. This approach is implemented, for example, in the software R (R Development Core Team, 2010) in the package mblm (see Komsta, 2013) that includes many tools of Theil’s regression. The confidence interval based on the Wilcoxon signed rank test has been ascribed to John Tukey (see Hollander & Wolfe, 1999 for historical details). However, it was originally developed to obtain a confidence interval for the true center of symmetry of a distribution from which we observed a sample of iid data, whereas it turns out very quickly that in case of slope estimation in Theil’s regression the input data are definitely not independent. Therefore, the true confidence level of the resulting interval is of question and our paper aims to study this issue.

The paper is organized as follows. Section 2 defines the underlying model. In Sections 3 and 4 we provide detailed description of the classical Theil’s confidence interval and the confidence interval based on the Tukey’s methodology, respectively. Section 5 shows by means of Monte Carlo simulations that the latter has its true confidence level under the nominal confidence level which is set to the traditional 95% throughout the whole paper. We prove this observation rigorously in case of $n = 4$ data points (Section 7). In Section 6 we provide a proof also in case of $n = 5$ data points; here, its final part requires uniform distribution of the random errors in order to numerically evaluate the true confidence level of the interval discussed. For the sake of completeness we treat also the case of $n = 3$ data points in Section 8. Finally, in Section 9 we briefly discuss the R implementation of the confidence interval for the true slope based on the Tukey’s methodology. Proofs of the theorems were deferred to the Appendix.

2 The model
At each of the $n$ fixed points $x_1, x_2, \ldots, x_n$ (values of the predictor $x$) we observe the value of a random variable $Y$ (response). We get a set of observations $Y_1, Y_2, \ldots, Y_n$, where $Y_i$ is the response at $x_i$. Without loss of generality we assume that $x_1 < x_2 < \cdots < x_n$. Our linear model has the form

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\beta_0$ (intercept) and $\beta_1$ (slope) are unknown parameters. Finally, the unobservable random errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are iid random variables from a continuous distribution.

*Corresponding author: somorci@fmph.uniba.sk
3 Theil’s confidence interval for slope

The hypothesis

\[ H_0 : \beta_1 = \beta^* \]

can be tested using \( D_i = Y_i - \beta^* x_i = (\beta_1 - \beta^*) x_i + \beta_0 + \varepsilon_i \). Provided that \( H_0 \) is true, the \( D_i \)'s do not depend on the \( x_i \)'s, i.e. they do not correlate. Hence, the validity of \( H_0 \) can be “measured” e.g. by the sample Kendall’s correlation coefficient

\[ \tau = \frac{N_c - N_d}{\binom{n}{2}}, \]

where \( N_c \) is the number of concordant pairs (i.e. pairs of points \([x_i, D_i]\) and \([x_j, D_j]\) such that \((x_i - x_j)(Y_i - Y_j) > 0\)) and \( N_d \) is the number of discordant pairs (i.e. pairs of points \([x_i, D_i]\) and \([x_j, D_j]\) such that \((x_i - x_j)(D_i - D_j) < 0\)). The test statistic \( K = \binom{n}{2} \tau = N_c - N_d \) is known as the Kendall \( K \) statistic (see Hollander & Wolfe, 1999). By \( K_n \) we denote its distribution under independence of the \( D_i \)'s from the \( x_i \)'s. The distribution \( K_n \) has been tabulated (see e.g. Hollander & Wolfe, 1999) and implemented in many statistical softwares (see e.g. Wheeler, 2009), because it depends just on the sample size \( n \), but not on the distribution of the data. The distribution \( K_n \) is discrete, symmetric and has the support

\[ \{-\binom{n}{2}, -\binom{n}{2} + 2, -\binom{n}{2} + 4, \ldots, \binom{n}{2} - 2, \binom{n}{2}\}, \]

because \( K \) has the same parity as \( \binom{n}{2} \). A test of the hypothesis \( H_0 \) at the significance level \( \alpha \) is then

\[ \text{reject } H_0 : \beta_1 = \beta^*, \text{ if } |K| \geq k_n(\alpha/2), \]

where \( k_n(\alpha/2) \) stands for the upper quantile of the distribution \( K_n \). It should be such an integer that, under \( H_0 \), \( P(K \geq k_n(\alpha/2)) = \alpha/2 \). However, due to the discrete nature of the distribution \( K_n \), an exact equality is virtually impossible. Therefore, we define \( k_n(\alpha/2) \) as such a unique integer with the same parity as \( \binom{n}{2} \) that

\[ P(K \geq k_n(\alpha/2)) \leq \alpha/2 \quad \text{and} \quad P(K \geq k_n(\alpha/2) - 2) > \alpha/2. \]

The consequence is that in general the true probability of the type I error of the above test is bellow the nominal significance level \( \alpha \), because it equals \( 2 \cdot P(K \geq k_n(\alpha/2)) \).

Similar idea leads to a confidence interval for the true slope \( \beta_1 \). Denote

\[ S_{ij} = \frac{Y_i - Y_j}{x_i - x_j}, \quad (i < j) \]

the “sample” slope of the line given by the pair of sample points \([x_i, Y_i]\) and \([x_j, Y_j]\). There are \( N = \binom{n}{2} = n(n-1)/2 \) such slopes. Order them ascendingly and denote the resulting sequence \( s_1 < s_2 < \cdots < s_N \) — since the random \( \varepsilon_i \)'s come from a continuous distribution, we may ignore ties between the sample slopes, because they will happen with zero probability, i.e. we shall assume that there are sharp inequalities between the \( s_i \)'s. The Theil’s \( 1 - \alpha \) confidence interval for the true slope \( \beta_1 \) is

\[ (s_l, s_u), \quad (1) \]

where

\[ l = \frac{N - k_n(\alpha/2)}{2} + 1 \quad \text{and} \quad u = \frac{N + k_n(\alpha/2)}{2}; \]

see e.g. Hollander and Wolfe (1999). Note that the indices \( l \) and \( u \) are symmetric in the sense that the same confidence interval is obtained also by taking the \( l \)-th slope from below a the \( l \)-th slope from above, since \( l + u = N + 1 \). The discrete nature of the distribution \( K_n \) involved causes the true confidence level of the above interval to be typically over \( 1 - \alpha \). The exact value is given by the following theorem.

**Theorem 1.** For all \( k \) of the form \( N - 2i \) \((i = 0, 1, \ldots, \lfloor N/4 \rfloor)\) put \( l = (N - k)/2 + 1 \) and \( u = (N + k)/2 \). Then the true confidence level of the Theil-type confidence interval \((s_l, s_u)\) is \( 1 - 2 \cdot P(K \geq k) \) where \( K \) is a random variable following \( K_n \).

Theorem 1 implies that the true confidence level of the Theil’s confidence interval (1) equals \( 1 - 2 \cdot P(K \geq k_n(\alpha/2)) \), which is at least \( 1 - \alpha \).
4 An à la Tukey confidence interval for slope

We start with a brief description of the Tukey’s methodology in a general setting (see e.g. Hollander & Wolfe, 1999).

Suppose we have some input iid random variables $Z_1, Z_2, \ldots, Z_N$ coming from a continuous symmetric distribution. We compute the so-called Walsh averages $(Z_i + Z_j)/2$ ($i \leq j$). Now we order the Walsh averages ascendingly (due to the continuity of the underlying distribution, we may ignore ties) and denote the resulting set $w_1 < w_2 \cdots < w_P$, where $P = \binom{N}{2} + N = N(N + 1)/2$. Then the Tukey’s $1 - \alpha$ confidence interval for the center of symmetry of the true distribution of the $Z_i$’s will be

$$(w_L, w_U),$$

where $U = t_N(\alpha/2)$. The value of $L$ will be “symmetric” in the sense that $L = P - t_N(\alpha/2) + 1$, i.e. one takes the $t_N(\alpha/2)$-th Walsh average from bellow and the $t_N(\alpha/2)$-th from above. Finally, $t_N(\alpha/2)$ denotes the $\alpha/2$ upper quantile of the null distribution of the Wilcoxon’s signed rank statistic $T^+$ having the range $0, 1, 2, \ldots, P$ (see e.g. Hollander & Wolfe, 1999 for details). Similarly as with the Theil’s confidence interval, the true confidence level of the Tukey’s confidence interval is typically strictly above $1 - \alpha$.

Now we apply the Tukey’s methodology described above to obtain a confidence interval for the true slope. The role of the $Z_i$’s will be played by the set of the slopes of all lines given by all pairs of the data points, i.e. by the set $\{S_{ij}; i < j\}$. We name the resulting interval, i.e. the Tukey’s confidence interval based on the slopes $S_{ij}$, the à la Tukey confidence interval. Unfortunately, one of the basic assumptions of the Tukey’s methodology is independence of the input data, i.e. independence of the $Z_i$’s. However, it is easily seen, that this assumption does not hold for the slopes $S_{ij}$. Actually, there is functional dependence among the slopes, because, for example, the knowledge of $S_{1,1}, S_{1,2}, \ldots, S_{1,n-1}$ enables us to compute the remaining $S_{ij}$’s, because

$$S_{ij} = \frac{S_{ij}(x_i - x_j) - S_{ji}(x_i - x_j)}{x_i - x_j}.$$

This does not necessarily mean that the à la Tukey confidence interval does not provide at least the nominal confidence level. For example, it may happen that the nominal confidence level is preserved, just the interval is redundantly wide. The best scenario from the à la Tukey confidence interval’s point of view is that the interval provides the nominal level of confidence while being narrower than the classical Théil’s confidence interval described in Section 3.

5 Monte Carlo study

However, none of the scenarios described in the previous paragraph seems to be true, as can be seen from the results of a Monte Carlo study we have conducted. In Table 1 we provide simulation estimates of the true confidence levels of the à la Tukey confidence interval for the true slope under various settings.

The number of data points $n$ changed from 6 to 200. The true values of intercept $\beta_0$ and slope $\beta_1$ were set to 0 and 1, respectively. The iid random errors $\varepsilon_i$ were generated from the normal distribution $N(0, 0.01)$, the Cauchy distribution with location parameter 0 and scale parameter 0.1, or the uniform distribution on the interval $(-0.2, 0.2)$. The motivation for the scale parameters of the distributions was to make the spread of the $\varepsilon_i$’s comparable with the spread of the $x_i$’s, i.e. to achieve that the data points $(x_i, Y_i)$ do not produce an ideally straight line, nor resemble a shapeless data cloud. In the left part of Table 1, the $x_i$’s created an equidistant design on the interval $(0, 1)$, more precisely, $x_i = (i - 1)/(n - 1)$ for $i = 1, 2, \ldots, n$. In the right part of the table, the design of the experiment consisted of two clusters of evenly spaced points on the subintervals $(0, 1/3)$ and $(2/3, 1)$, more precisely, $x_i = (i - 1)/(3(n/2 - 1))$ for $i = 1, 2, \ldots, n/2$ and $x_i = 2/3 + (i - n/2 - 1)/(3(n/2 - 1))$ for $i = n/2 + 1, n/2 + 2, \ldots, n$. Each figure in Table 1 is based on 10,000 simulations and it is the proportion of times (rounded to three decimal places) the confidence interval covered the true slope $\beta_1$. The nominal confidence level of the à la Tukey confidence interval was set to 95%.

The main message of Table 1 is that, irrespective of the probability distribution of the random errors, the true confidence level of the à la Tukey confidence interval is strictly below the nominal 95% and decreases rapidly with increasing number of data $n$ in all settings presented in our study. The design of the $x_i$’s does not seem to play an important role either: we tried also some other designs not reported here and the resulting figures were very similar. The Monte Carlo simulations support our suspicion that
Two clusters of evenly spaced
Distribution of random errors:
normal Cauchy uniform
table

| Number of data points | Evenly spaced \( x_i \)'s | Distribution of random errors: | Two clusters of evenly spaced \( x_i \)'s | Distribution of random errors: |
|-----------------------|-----------------------------|--------------------------------|-----------------------------|--------------------------------|
|                       | normal Cauchy uniform       | normal Cauchy uniform         | normal Cauchy uniform       | normal Cauchy uniform         |
| 6                     | .869 .850 .867              | .871 .855 .867               | .869 .850 .867              | .871 .855 .867               |
| 10                    | .804 .773 .793              | .804 .777 .800               | .804 .773 .793              | .804 .777 .800               |
| 20                    | .679 .636 .675              | .678 .646 .675               | .679 .636 .675              | .678 .646 .675               |
| 30                    | .591 .551 .588              | .595 .561 .596               | .591 .551 .588              | .595 .561 .596               |
| 50                    | .533 .494 .540              | .530 .499 .541               | .533 .494 .540              | .530 .499 .541               |
| 100                   | .399 .365 .402              | .402 .373 .403               | .399 .365 .402              | .402 .373 .403               |
| 120                   | .329 .307 .329              | .338 .310 .334               | .329 .307 .329              | .338 .310 .334               |
| 140                   | .311 .290 .310              | .302 .293 .309               | .311 .290 .310              | .302 .293 .309               |
| 160                   | .294 .267 .295              | .303 .278 .296               | .294 .267 .295              | .303 .278 .296               |
| 180                   | .276 .246 .278              | .276 .251 .278               | .276 .246 .278              | .276 .251 .278               |
| 200                   | .261 .240 .265              | .268 .249 .266               | .261 .240 .265              | .268 .249 .266               |

6 The case of \( n = 5 \) data points

We are going to examine the true confidence level of the à la Tukey confidence interval under the assumption of the nominal confidence level \( 1 - \alpha = 95\% \). For \( n = 5 \), \( N = \binom{5}{2} = 10 \) and \( \alpha = 5\% \) the appropriate upper quantiles \( k_5(2.5\%) \) and \( t_{10}(2.5\%) \) needed for the constructions of the Theil’s and the à la Tukey confidence interval are 10 and 47, respectively (see e.g. Table A.30 and Table A.4 in Hollander and Wolfe (1999)). The resulting Theil’s confidence interval (1) and the à la Tukey confidence interval (2) are \( (s_1, s_{10}) \) and \( (w_9, w_{47}) \), respectively. It means that the former is given by the minimum and the maximum sample slope and the latter by the 9-th Walsh average from bellow and the 9-th Walsh average from above — which is the 47-th from bellow, since there are together \( P = 10 \cdot (10 + 1)/2 = 55 \) Walsh averages. The following theorem shows the relationship between \( (s_1, s_{10}) \) and \( (w_9, w_{47}) \).

**Theorem 2.** For \( n = 5 \), the 95% à la Tukey confidence interval \( (w_9, w_{47}) \) is always a subset of the 95% Theil’s confidence interval \( (s_1, s_{10}) \).

Theorem 2 itself is not enough to claim that the à la Tukey confidence interval has its true confidence level under 95%. However, by Monte Carlo simulations not reported here we noticed that the à la Tukey confidence interval \( (w_9, w_{47}) \) happens to be very often the subset of the even narrower Theil-type confidence interval \( (s_2, s_9) \). The true confidence level of \( (s_2, s_9) \) can be obtained easily: put \( l = 2 \), \( u = 9 \), \( n = 5 \), and \( N = 10 \), then the notation of Theorem 1 implies that \( k = 8 \) and the theorem itself gives the true confidence \( 1 - 2 \cdot P(K \geq 8) \) which can be evaluated e.g. by Table A.30 in Hollander and Wolfe (1999). The approximate result is 91.67%. Therefore, our aim is to show, that the à la Tukey confidence interval \( (w_9, w_{47}) \) is “very often” a subset of \( (s_2, s_9) \) with the poor confidence level of 91.67%. The consequence will be, that although the true confidence level of the à la Tukey confidence interval \( (w_9, w_{47}) \) could be over that of \( (s_2, s_9) \), it is definitely under 95%. The next theorem states exact conditions in terms of the \( s_i \)'s when the above-described desired “very often” inclusion of \( (w_9, w_{47}) \) in \( (s_2, s_9) \) happens, i.e. conditions when the lower (upper) bound of \( (w_9, w_{47}) \) is over (under) the lower (upper) bound of \( (s_2, s_9) \).

**Theorem 3.** If \( n = 5 \) then: a) The random event \( s_2 \leq w_9 \) occurs if and only if \( 2s_2 \leq s_1 + s_9 \). b) The random event \( w_{47} \leq s_9 \) occurs if and only if \( s_2 + s_{10} \leq 2s_9 \).

The following theorem provides an upper bound for the true confidence level \( P(\beta_1 \in (w_9, w_{47})) \) of the à la Tukey confidence interval \( (w_9, w_{47}) \).
Theorem 4. Let $n = 5$. Denote

$$p_1 = P(\beta_1 \in (s_2, s_9) \land 2s_2 \leq s_1 + s_9 \land s_2 + s_{10} \leq 2s_9)$$

$$p_2 = P(\beta_1 \in (s_2, s_{10}) \land 2s_2 \leq s_1 + s_9 \land 2s_9 < s_2 + s_{10})$$

$$p_3 = P(\beta_1 \in (s_1, s_9) \land s_1 + s_9 < 2s_2 \land s_2 + s_{10} \leq 2s_9)$$

$$p_4 = P(\beta_1 \in (s_1, s_{10}) \land s_1 + s_9 < 2s_2 \land 2s_9 < s_2 + s_{10})$$

Then $P(\beta_1 \in (w_9, w_{47})) \leq p_1 + p_2 + p_3 + p_4$.

Let us discuss Theorem 4 in greater detail. The probability $p_1$ is simply the true confidence level of $(s_2, s_9)$ together with the probability of the above-discussed “very often” inclusion of $(w_9, w_{47})$ in $(s_2, s_9)$ with the poor true confidence level 91.67%. This means that $p_1$ is at most 91.67%. The probabilities $p_2$, $p_3$, $p_4$ are based on the true confidence levels of the slightly wider intervals $(s_2, s_{10})$, $(s_1, s_9)$, $(s_1, s_{10})$ and these confidence levels can be high. However, from the proof of Theorem 4 in the Appendix it can be seen that the probabilities $p_2$, $p_3$, and $p_4$ reflect only cases when these intervals include also $(w_9, w_{47})$ whereas $(w_9, w_{47})$ is not completely included in $(s_2, s_9)$. Monte Carlo simulations suggest that these cases are rare, therefore we hope that $p_2$, $p_3$, and $p_4$ turn out to be low.

All we need to do is to determine the probabilities $p_1$, $p_2$, $p_3$, and $p_4$. Note that they are given just by the joint probability distribution of the ordered slopes $s_i$, i.e. one does not need to examine the much more complicated joint probability distribution of the ordered Walsh averages $w_i$ anymore. Still, it is not easy to evaluate $p_1$, $p_2$, $p_3$, $p_4$ under arbitrary probability distribution of the random errors $\varepsilon_i$.

Therefore, at this point we introduce an additional condition that makes the evaluation easier:

**Condition 1.** The probability distribution of the error terms $\varepsilon_i$ is uniform on the interval $(-1, 1)$.

There are $10!$ possible orderings of the ten slopes $S_{1,2}, S_{1,3}, \ldots, S_{4,5}$. Denote $B_1, B_2, \ldots, B_{10!}$ the appropriate random events, i.e. each $B_i$ denotes an event that a particular ordering of the slopes happened. Then $p_1$ can be decomposed as

$$p_1 = \sum_{i=1}^{10!} P(\{\beta_1 \in (s_2, s_9) \land 2s_2 \leq s_1 + s_9 \land s_2 + s_{10} \leq 2s_9\} \cap B_i).$$

Let $B_i$ denote, for example, the ordering

$$S_{4,5} < S_{3,5} < S_{2,5} < S_{1,5} < S_{2,3} < S_{2,4} < S_{1,3} < S_{1,4} < S_{3,4} < S_{1,2}.$$  \hspace{1cm} (3)

Since

$$S_{ij} = \frac{Y_i - Y_j}{x_i - x_j} = \beta_i + \frac{\varepsilon_i - \varepsilon_j}{x_i - x_j},$$ \hspace{1cm} (4)

the 9 inequalities in (3) that define the ordering can be rewritten as 9 linear inequalities of the form

$$c_1\varepsilon_1 + c_2\varepsilon_2 + \cdots + c_5\varepsilon_5 < c_6,$$ \hspace{1cm} (5)

where $c_0, c_1, c_2, \ldots, c_5$ are constants depending on the $x_i$’s. Further, under $B_i$ the conditions $\beta_1 \in (s_2, s_9)$, $2s_2 \leq s_1 + s_9$ and $2s_9 \geq s_2 + s_{10}$ can be also rewritten as $2+1+1=4$ inequalities of the form (5), because under $B_i$, for example, $s_1 = \beta_1 + \frac{\varepsilon_1 - \varepsilon_2}{x_4 - x_5}$, $s_2 = \beta_1 + \frac{\varepsilon_2 - \varepsilon_3}{x_3 - x_5}$, etc. Note that under Condition 1 the probability distribution of the vector $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_5)^T$ is uniform in the 5-dimensional cube with the vertices $[\pm1, \pm1, \pm1, \pm1, \pm1]$ and edges of length 2. Therefore,

$$p_1 = \sum_{i=1}^{10!} \frac{V_i}{25},$$

where $V_i$ is the volume of the 5-dimensional polytope with faces given by the above-mentioned $9+4=13$ linear inequalities of the form (5) and 10 additional inequalities

$$\varepsilon_i < 1 \text{ and } \varepsilon_i > -1 \quad (i = 1, 2, \ldots, 5)$$

defining the boundary of the cube. To evaluate the volume of a polytope given by the defining $13+10=23$ inequalities of the form (5) we used the specialized software Vinci (Büeler & Enge, 2003). The 23 inequalities defining the polytope are passed to Vinci and the computation of the volume is then based on the triangulation of the polytope and computation of determinants.

Nevertheless, we still have to evaluate volumes of $10! = 3,628,800$ polytopes and evaluation of $p_2$, $p_3$, and $p_4$ is going to quadruple this number. Fortunately, most of these polytopes have zero volume because the following theorem implies that a lot of the $10!$ possible orderings of the slopes are impossible.
Theorem 5. Let \( a < b < c \) be three indices. Then the slope \( S_{ac} \) is neither the greatest nor the smallest of the trio \( S_{ab}, S_{ac}, \) and \( S_{bc} \).

An automatized computer inspection of the 10! possible orderings revealed quickly, that only 768 of them conform to Theorem 5. At this point we have to set concrete values of the \( x_i \)'s, because the \( p_i \)'s depend on them. We decided for the following.

Condition 2. The \( x_i \)'s create an equidistant design \( x_i = i \) for \( i = 1, 2, \ldots, n \).

As a byproduct, this choice of the \( x_i \)'s has the following pleasant consequences that reduce the amount of computations.

Theorem 6. If Condition 2 holds and \( n = 5 \), then \( p_4 = 0 \).

Theorem 7. If Conditions 1 and 2 hold, then \( p_2 = p_3 \).

Hence, it suffices to evaluate \( p_1 \) and \( p_2 \) using Vinci as the volumes of the above-mentioned 768 + 768 polytopes and by Theorems 6 and 7 one obtains

\[
p_1 + p_2 + p_3 + p_4 = 0.8107315 + 0.0595787 + 0.0595787 + 0 = 92.98889%. 
\]

This provides an upper bound for the true confidence level of the interval \((w_9, w_{47})\), i.e. the true confidence level is under the nominal level 95\% (as a Monte Carlo estimate based on 1,000,000 simulations we obtained 87.9\%). We note that Condition 2 is just technical, since we are able to evaluate \( p_1 + p_2 + p_3 + p_4 \) under any particular arrangement of the \( x_i \)'s. However, Condition 1 about the uniform distribution of the random errors is crucial, because it reduced our computation to evaluation of volumes of polytopes, which could be accomplished by Vinci.

7 The case of \( n = 4 \) data points

In case of \( n = 4 \) data points, the 95\% à la Tukey confidence interval will be \((w_1, w_{21})\) which is obviously the same as \((s_1, s_6)\). Its true confidence level can be evaluated by Theorem 1: put \( l = 1, \ u = 6, \ n = 4, \ N = 6 \), obtain \( k = 6 \) and the theorem gives the confidence \( 1 - 2 \cdot P(K \geq 6) \), which can be evaluated e.g. by Table A.30 in Hollander and Wolfe (1999). The approximate result is 91.67\% which is definitely under 95\%, i.e. the à la Tukey confidence interval does not work correctly in this case either. Note that — unlike the case of \( n = 5 \) data points — the obtained result 91.67\% holds in general, e.g. it is completely independent of the additional Conditions 1 or 2.

The above paragraph also means that in case of \( n = 4 \) data points the Theil’s approach is unable to produce a 95\% confidence interval, because the confidence level of \((s_1, s_6)\) (the widest Theil-type interval) is under 95\%. From another point of view, the 95\% Theil’s confidence interval cannot be produced, because \( k_4(2.5\%) \) satisfying our definition of the upper quantile value does not exist.

8 The case of \( n = 3 \) data points

With just \( n = 3 \) data points at hand, the Theil’s approach breaks down, because \( k_3(2.5\%) \) does not exist. The same happens to the 95\% à la Tukey confidence interval, because the Tukey’s methodology does not work for such a low number of data and nominal confidence level of 95\% (\( t_3(2.5\%) \) does not exist).

9 An R implementation of the à la Tukey confidence interval

As we already noted, the à la Tukey confidence interval for the true slope is implemented e.g. in the R package nblm that is available in the CRAN package repository since 2005. Since that time the package documentation has been noting that the package does not implement the original Theil’s confidence interval based on Kendall’s tau but it is considered to be implemented in next version of the package. However, it has not been implemented till now (February 2015), despite the fact that already the third version of the package has been released. Nevertheless, the main problem is that the package does not provide any warning about the deflated true confidence level of the intervals produced. The only exceptions are the cases of \( n = 4 \) and \( n = 3 \) data points. In case of \( n = 4 \) data points the package nblm produces a correct warning message, that the requested confidence level is not achievable. However,
careful inspection of the package code reveals that it is just a coincidence: in fact, the warning says nothing about the confidence interval for the true slope, because it has been invoked by the computation of a confidence interval for the true intercept (this confidence interval was not discussed in our paper). For \( n = 3 \) data points the package mblm produces an error message, however, as before the true reason for the message is a problem with the computation of a confidence interval for the true intercept.

10 Conclusions

We have shown by means of Monte Carlo simulations that the à la Tukey confidence interval for the true slope in the straight line regression model seems to be unable to achieve the nominal confidence level. The loss of interval’s confidence does not seem to depend too much on the design of the experiment or on the distribution of the random error errors, but becomes very serious with increasing number of data — in all cases with over 160 data points we observed the true confidence level even under 30\% instead of the nominal 95\%.

In case of \( n = 4 \) data points we easily obtained also the true confidence level of the à la Tukey confidence interval — the simplicity of the reasoning resulted from the fact that the lower and upper limit of the à la Tukey confidence interval turned out to be some of the original sample slopes. However, in case of \( n = 5 \) data points the situation was much more complicated: we were able to obtain only an upper bound for the true confidence level and we numerically evaluated this upper bound under the condition of uniformly distributed random errors.

Theoretically, the process of evaluation of the probabilities \( p_i \) appearing in the above mentioned upper bound can be adopted to obtain the exact value of the true confidence level of the à la Tukey interval. However, already in case of \( n = 5 \) data points there are 55 Walsh averages given by the ten slopes \( S_{ij} \) and, theoretically, these Walsh averages can be arranged in 55! permutations. These would result in the necessity to evaluate and sum volumes of as much as 55! \( \approx 1.27 \cdot 10^{73} \) polytopes — a very hard task from the numerical point of view. Similarly as in the evaluation of the \( p_i \)'s, many of these polytopes could be a priori shown to be of zero volume, but we decided to proceed in a different way: we estimated the true confidence level from above by terms not involving the Walsh averages and showed rather easily that this upper bound is strictly under 95\%.

A natural question arises, if the reasoning in case of \( n = 5 \) data points can be easily adopted or even generalized for larger \( n \). Despite our effort we have not found any positive answer, because the situation complicates dramatically already for \( n = 6 \).

The à la Tukey confidence interval for the true slope is implemented in the R package mblm without any warning about its deflated true confidence level. The results of our paper show that this functionality of the package (i.e. computation of the confidence interval for the true slope) should not be used, because it tends to provide too liberal interval estimates.

Acknowledgments

The work was supported by the VEGA grant No. 2/0047/15 of the Science Grant Agency of the Slovak Republic.

References

Büeler, B., & Enge, A. (2003). Vinci [Computer software manual]. Retrieved from http://www.math.u-bordeaux1.fr/~aenge/index.php?category=software&page=vinci (version 1.0.5)
Hollander, M., & Wolfe, D. A. (1999). Nonparametric statistical methods – 2nd ed. New York: John Wiley & Sons.
Komsta, L. (2013). mblm: Median-based linear models [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=mblm (R package version 0.12)
R Development Core Team. (2010). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org (ISBN 3-900051-07-0)
Theil, H. (1950a). A rank-invariant method of linear and polynomial regression analysis, I. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, 53, 386–392.
Appendix

Proof of Theorem 1
Recall $N_c, N_d$ and the hypothesis $H_0$ from Section 3. In Theil (1950a) on p. 390 the true confidence level of the Theil-type confidence interval $(s_i, s_u)$ is expressed as

$$1 - 2 \cdot P(N_d \leq l - 1),$$

where the probability is evaluated under $H_0$ and the result holds even under a more general setting than discussed in our paper. Since $K = N_c - N_d$ and $N_c + N_d = N$, we obtain that $N_d = (N - K)/2$ and

$$1 - 2 \cdot P(N_d \leq l - 1) = 1 - 2 \cdot P\left(\frac{N - K}{2} \leq \left(\frac{N - k}{2} + 1\right) - 1\right) = 1 - 2 \cdot P(K \geq k).$$

\[\square\]

Proof of Theorem 2
Since the smallest Walsh average $w_1$ is given by the smallest slope $s_1$ as $(s_1 + s_1)/2 = s_1$ and the largest Walsh average $w_{55}$ is given by the largest slope $s_{10}$ as $(s_{10} + s_{10})/2 = s_{10}$, we obtain $s_1 = w_1 < w_9$ and $w_{47} < w_{55} = s_{10}$. \[\square\]

Proof of Theorem 3

Part a): We prove the equivalent statement “$w_9 < s_2$ iff $s_1 + s_9 < 2s_2$.” Start with $w_9 < s_2$ and consider the 9 smallest Walsh averages $w_1 < w_2 < \cdots < w_9$. Each of them is of the form $(s_i + s_j)/2$ for some $i \leq j$ and since $s_2 = (s_2 + s_2)/2$, the assumption $w_9 < s_2$ means that

$$\frac{s_i + s_j}{2} < \frac{s_2 + s_2}{2}.$$  \hspace{1cm} (6)

Because $s_1 < s_2 < \cdots < s_{10}$, the sharp inequality (6) immediately implies, that $i = 1$ and the 9 smallest Walsh averages $w_1 < w_2 < \cdots < w_9$ have to be of the form $(s_1 + s_1)/2 < (s_1 + s_2)/2 < \cdots < (s_1 + s_9)/2$. Therefore, the inequality $w_9 < s_2$ can be rewritten as $(s_1 + s_9)/2 < (s_2 + s_2)/2$ and the first part of the proof is complete.

Now, start with $s_1 + s_9 < 2s_2$, i.e. $(s_1 + s_9)/2 < s_2$. Since the 8 Walsh averages $(s_1 + s_1)/2 < (s_1 + s_2)/2 < \cdots < (s_1 + s_8)/2$ are even smaller then $(s_1 + s_9)/2$, we see that there are at least 9 Walsh averages smaller than $s_2$. Therefore, also the 9-th smallest Walsh averages, i.e. $w_9$, is smaller than $s_2$. \[\square\]

Part b): Note that the proof of part a) is based on the natural ordering “the higher slope (or Walsh average), the higher index”. Using the reverse ordering “the higher slope (or Walsh average), the lower index” in the proof of part a), one obtains the “symmetric” counterpart of part a), which is part b). \[\square\]

Proof of Theorem 4
Split the whole probability space into these four disjoint random events:

$$A : s_2 \leq w_9 \land w_{47} \leq s_9$$
$$B : s_2 \leq w_9 \land s_9 < w_{47}$$
$$C : w_9 < s_2 \land w_{47} \leq s_9$$
$$D : w_9 < s_2 \land s_9 < w_{47}$$

Denote by $U$ the random event $\{\beta_1 \in (w_9, w_{47})\}$. Note that the minimum and the maximum of all slopes $s_i$ and their Walsh averages $w_i$ are $s_1$ and $s_{10}$, respectively. This implies that

$$P(U \cap A) \leq P(\{\beta_1 \in (s_2, s_9)\} \cap A) = p_1,$$

$$P(U \cap B) \leq P(\{\beta_1 \in (s_2, s_{10})\} \cap B) = p_2,$$

$$P(U \cap C) \leq P(\{\beta_1 \in (s_1, s_9)\} \cap C) = p_3,$$

$$P(U \cap D) \leq P(\{\beta_1 \in (s_1, s_{10})\} \cap D) = p_4,$$
where the final equality in each row follows from Theorem 3. Hence, we obtain
\[ P(U) = P(U \cap A) + P(U \cap B) + P(U \cap C) + P(U \cap D) \leq p_1 + p_2 + p_3 + p_4. \]

\[ \square \]

**Proof of Theorem 5**

By contradiction, let \( S_{ac} \) be the greatest of \( S_{ab}, S_{ac}, S_{bc} \) — the case that \( S_{ac} \) is the smallest can be treated analogously. By (4) and by noting that \( x_a < x_b < x_c \), one observes that the inequality \( S_{ac} > S_{ab} \) is equivalent to
\[ (\varepsilon_a - \varepsilon_c)(x_a - x_b) > (\varepsilon_a - \varepsilon_b)(x_a - x_c) \]
and \( S_{ac} > S_{bc} \) is equivalent to
\[ (\varepsilon_a - \varepsilon_c)(x_b - x_c) > (\varepsilon_b - \varepsilon_c)(x_a - x_c) \].

By summing these two inequalities we obtain \((\varepsilon_a - \varepsilon_c)(x_a - x_c) > (\varepsilon_a - \varepsilon_c)(x_a - x_c)\) which is impossible. \[ \square \]

**Proof of Theorem 6**

Theorem 5 implies that the minimum and maximum sample slopes \( s_1 \) and \( s_{10} \) are of the form \( s_1 = S_{i,i+1} \) and \( s_{10} = S_{j,j+1} \) for some distinct \( i \) and \( j \) from \( \{1, 2, 3, 4\} \). Straightforward algebra implies that
\[ S_{i,j} - S_{i+1,j+1} = (s_{10} - s_1)/(i-j) \] under Condition 2, which means that
\[ s_{10} = S_{i+1,i+2} \]

Note that \(|i-j| \leq 3 \) (because \( 1 \leq i, j \leq 4 \)) and if both \( S_{i,j} \) and \( S_{i+1,j+1} \) belong to \( \{s_2, s_3, \ldots, s_9\} \), then one obtains from (7) that
\[ s_{10} - s_1 \leq 3(s_9 - s_2). \]

However, summing the inequalities \( s_1 + s_9 < 2s_2 \) and \( 2s_9 < s_2 + s_{10} \) appearing in the definition of \( p_4 \) yields
\[ s_{10} - s_1 > 3(s_9 - s_2), \]
which contradicts (8), i.e. \( p_4 = 0 \).

It remains to treat the case when not both \( S_{i,j} \) and \( S_{i+1,j+1} \) belong to \( \{s_2, s_3, \ldots, s_9\} \). This happens if and only if \(|i-j| = 1 \). Without loss of generality, we will suppose that \( j = i + 1 \), i.e. \( s_1 = S_{i,i+1}, s_{10} = S_{i+1,i+2} \) and \( i \leq 3 \).

a) The case when \( i \leq 2 \). We will show that the inequality
\[ 2s_9 < s_2 + s_{10} \] appearing in the definition of \( p_4 \) is impossible. Because \( s_2 \leq S_{i+2,i+3} \leq s_9 \) and \( s_{10} = S_{i+1,i+2} \), the inequality (9) would imply that \( 2S_{i+1,i+3} < S_{i+2,i+3} + S_{i+1,i+2} \), which is equivalent to \( 0 < 0 \) under Condition 2.

b) The case when \( i = 3 \). We will show that the inequality
\[ s_1 + s_9 < 2s_2 \] appearing in the definition of \( p_4 \) is impossible. Because \( s_2 \leq S_{2,4}, S_{2,3} \leq s_9 \) and \( s_1 = S_{3,4} \), the inequality (10) would imply that \( S_{3,4} + S_{2,3} < 2S_{2,4} \), which is equivalent to \( 0 < 0 \) under Condition 2. \[ \square \]

**Proof of Theorem 7**

Symmetry and independence of the distribution of the \( \varepsilon_i \)'s given by Condition 1, together with the equidistantness of the \( x_i \)'s given by Condition 2 means that moving from the \( \varepsilon_i \)'s to the “equiprobable” \(-\varepsilon_i \)'s reverts the ordering of the sample slopes and also the ordering of their Walsh averages, because each sample slope changes symmetrically around \( \beta_1 \) (cf. (4)). It means that, for example, the sample slope with the label \( s_2 \) gets the label \( s_9 \), or the Walsh average with the label \( w_9 \) gets \( w_{47} \), etc. The relationships between the \( s_i \)'s and \( w_j \)'s change accordingly: for example, \( s_2 \leq w_9 \) changes to \( w_{47} \leq s_9 \).

Hence, we observe that the conditions defining \( p_2 \) change to conditions defining \( p_3 \). \[ \square \]