Consistency and Asymptotic Distribution of the Theil-Sen Estimator

Hanxiang Peng [a], Shaoli Wang [b], Xueqin Wang [b,*]

[a] *University of Mississippi, Department of Mathematics, University, MS 38677-1848*

[b] *Yale University, Department of Epidemiology and Public Health, New Haven, CT 06520-8034*

Abstract

In this paper, we obtain the strong consistency and asymptotic distribution of the Theil-Sen estimator in simple linear regression models with arbitrary error distributions. We show that the Theil-Sen estimator is super-efficient when the error distribution is discontinuous and that its asymptotic distribution may or may not be normal when the error distribution is continuous. We give an example in which the Theil-Sen estimator is not asymptotically normal. A small simulation study is conducted to confirm the super-efficiency and the non-normality of the asymptotic distribution.

Key words: Asymptotic normality; Linear regression; Median; Robustness; Super-efficiency

1 Introduction

We consider a simple linear regression model

$$Y_i = \beta x_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

* Corresponding author.
Email address: Xueqin.Wang@yale.edu (Xueqin Wang).

Preprint submitted to Elsevier Science *24 June 2007*

where x_i are nonidentical constants and ϵ_i independent and identically distributed (iid) random errors with an unknown cumulative distribution function (cdf) F. A well-known robust estimator of the slope β is the Theil-Sen estimator that was first proposed by Theil (1950) and then extended by Sen (1968). More precisely, we define

$$\mathcal{B}_n = \left\{ b_{ij} : b_{ij} = \frac{Y_j - Y_i}{x_j - x_i}, \text{ if } x_i \neq x_j, \ 1 \leq i < j \leq n \right\}. \tag{1.2}$$

Then the Theil-Sen estimator $\tilde{\beta}_n$ is defined as the median of all slopes in $\mathcal{B}_n$, $\tilde{\beta}_n = \text{med}(\mathcal{B}_n)$, where "med" stands for "median".

We deliberately leave out the intercept in model (1.1). Nonetheless, our model covers the linear regression model with an unknown intercept

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

by simply letting $\epsilon_i = \alpha + \varepsilon_i$, where ε_i is an error satisfying certain identifiability conditions. Our formulation in Model (1.1) does not impose any assumptions about the error. The intercept α can be estimated, for example, using the median of $\left\{ Y_i - \tilde{\beta}_n X_i : i = 1, \ldots, n \right\}$ under the identifiability condition that the error has a unique median. For the regression model with a zero intercept, a more robust estimator of the slope β is the median of the slopes of lines joining the origin with all observations (x_i, Y_i). This is the least absolute deviation estimator of the slope, which has a bounded influence function and a large-sample high breakdown point of 0.5. Since our principal focus in this paper is the asymptotic behavior of the slope estimator $\tilde{\beta}_n$ for the general regression model (1.1), we stop here the discussion of intercept estimations and the linear regression model through the origin.

The Theil-Sen estimator is robust with a high breakdown point of about 0.293 and also has a bounded influence function. It compares favorably with the ordinary least squares estimator in small-sample efficiency (Wilcox, 1998) and is competitive in terms of mean squared error with alternative slope estimators (Dietz, 1987).

The univariate Theil-Sen estimator has numerous multivariate extensions. Oja and Niinimaa (1984) generalized the Theil-Sen estimator to multiple regression models using pseudo observations and the Oja's median (1983). The Oja's median is a special spatial median. For the asymptotic properties of spatial medians, see Arcones et al. (1994) and Bose (1998). Zhou and Serfling (2006) gave another natural extension of the Theil-Sen estimator based on multivariate spatial U-quantiles. It is interesting to establish some of the properties of the univariate Theil-Sen estimator for its multivariate spatial extensions in multiple regression models, including semi-parametric generalized linear models, partially linear models and single index models. We pursue this matter in a separate study.

In review of the asymptotic results of the Theil-Sen estimator in the literature, we found that further study on this classical estimator was worthwhile. For instance, the consistency of the estimator has, to our knowledge, not been studied thus far. Sen (1968) investigated the asymptotic normality of the estimator only for *absolutely continuous* cdf F. However, as we point out in this paper, there is a gap in Sen's proof. Sen used a theorem from Hoeffding(1948), but his set-up does not satisfy the assumptions of the theorem.

In this paper we establish the strong consistency and asymptotic distribution of the Theil-Sen estimator for a general error distribution F (i.e., the cdf F of the error ϵ is arbitrary, thus including both *discontinuous and continuous* ones). To our surprise, the Theil-Sen estimator turns out to be *super-efficient* for discontinuous error distributions (See Section 2). We also obtain a general theorem on the asymptotic distribution (Theorem 3) when the error distribution is continuous (not necessarily absolutely continuous). The asymptotic normality claimed by Sen (1968) follows as a special case. We find that the Theil-Sen estimator is *not asymptotically normal* in general, though it does converge in distribution. We also give the conditions under which it is asymptotically normal (Remark 2 in Section 3). We provide two sets of conditions under which the general theorem holds. These conditions are easy to verify and satisfied by most common distributions; furthermore, they enable us to obtain an explicit formula for the scaling constant. Under these conditions, we show that the asymptotic distribution is normal when the cdf F is absolutely continuous and may *not be normal* when the cdf F is *not absolutely continuous.* An example is given in which the Theil-Sen estimator has a non-normal asymptotic distribution. We conduct a small simulation study that confirms the super-efficiency and the asymptotic non-normality.

The Theil-Sen estimator has been widely acknowledged in several popular textbooks on nonparametric statistics and robust regression. See, e.g., Sprent (1993), Hollander and Wolfe (1973 and 1999), and Rousseeuw and Leroy (2003). It also has been extensively studied in the literature. Sen (1968) and Wilcox (1998) investigated its asymptotic relative efficiency to the least squares estimator. Akritas et al. (1995) applied it to astronomy and Fernandes and Leblanc (2005) to remote sensing. Wang (2005) studied its asymptotic properties for Model (1.1) with a random covariate. Many of its extensions can be found in the literature, for example, in censored data; for details, see, e.g., Akritas et al. (1995), Jones (1997), and Mount and Netanyahu (2001).

The rest of this paper is organized as follows: In Section 2, we investigate consistency. In Section 3, we address asymptotic normality, present an example and conduct a small simulation study. In Section 4, we prove Theorem 3. Some technical details are given in the appendix.

2 Strong Consistency

In this section, we establish the strong consistency of the Theil-Sen estimator for both discontinuous and continuous error distributions. We start by introducing a general lemma.

First, for each $0 < r \leq \infty$, we divide the slope set $\mathcal{B}_n$ defined in (1.2) into two subsets $\mathcal{B}_{n,r}^+$ and $\mathcal{B}_{n,r}^-$:

$$\mathcal{B}_{n,r}^+ = \{b_{ij} \in \mathcal{B}_n : b_{ij} > \beta + 1/r\}, \qquad \mathcal{B}_{n,r}^- = \{b_{ij} \in \mathcal{B}_n : b_{ij} < \beta - 1/r\},$$

and define $N^+_{n,r} = \#(\mathcal{B}^+_{n,r})$ and $N^-_{n,r} = \#(\mathcal{B}^-_{n,r})$ as the cardinalities of $\mathcal{B}^+_{n,r}$ and $\mathcal{B}^-_{n,r}$, respectively.

Under Model (1.1), we can write $b_{ij} = \beta + e_{ij}$, $e_{ij} = (\epsilon_i - \epsilon_j)/(x_i - x_j)$, $x_i \neq x_j$. Then we have two U-statistics

$$\bar{N}^+_{n,r} = \frac{2}{n(n-1)} N^+_{n,r} = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \psi^+_r(Z_i, Z_j) \quad \text{and}$$
$$\bar{N}^-_{n,r} = \frac{2}{n(n-1)} N^-_{n,r} = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} \psi^-_r(Z_i, Z_j),$$

with kernels given by

$$\psi^+_r(Z_i, Z_j) = \mathbf{1}[e_{ij} > 1/r] \quad \text{and} \quad \psi^-_r(Z_i, Z_j) = \mathbf{1}[e_{ij} < -1/r],$$

respectively, where $\mathbf{1}[A]$ denotes the indicator of event A and $Z_i = (x_i, \epsilon_i)$.

Noticing that ϵ_i and ϵ_j are iid, we have

$$\begin{aligned} &E\psi^-_r(Z_i, Z_j) = E\psi^+_r(Z_i, Z_j) = P((\epsilon_i - \epsilon_j)/(x_i - x_j) > 1/r) \\ &= P(\epsilon_i - \epsilon_j > (x_i - x_j)/r)\mathbf{1}[x_i > x_j] + P(\epsilon_i - \epsilon_j < (x_i - x_j)/r)\mathbf{1}[x_i < x_j] \\ &= \mathbf{1}[x_i \neq x_j](1 - F_2(|x_i - x_j|/r)), \quad i < j, \end{aligned}$$

here F_2 is the cdf of $\epsilon_i - \epsilon_j$.

Now let a_n be the number of unequal pairs (x_i, x_j), namely, $a_n = \sum_{1\le i<j\le n} \mathbf{1}[x_i \neq x_j]$; and let $\bar{a}_n = 2a_n/n(n-1)$. Define

$$q_{n,r} = \frac{1}{a_n} \sum_{1\le i<j\le n} \mathbf{1}[x_i \neq x_j](1 - F_2(|x_i - x_j|/r)). \tag{2.1}$$

Then

$$E\{\bar{N}^-_{n,r}\} = E\{\bar{N}^+_{n,r}\} = \bar{a}_n q_{n,r}. \tag{2.2}$$

The next lemma gives the asymptotic behavior of the U-statistics, and its proof is put off to the last section.

Lemma 1 *Suppose that the nonrandom covariates $x_1, x_2, \ldots, x_n$ satisfy*

$$\frac{n^{-1}\log n}{\bar{a}_n^2} = o(1). \tag{2.3}$$

Then for $0 < r \leq \infty$,

$$\lim_{n\to\infty}\left(\bar{N}^+_{n,r}/\bar{a}_n - q_{n,r}\right) = 0, \quad \lim_{n\to\infty}\left(\bar{N}^-_{n,r}/\bar{a}_n - q_{n,r}\right) = 0, \quad \text{a.s.}$$

where a.s. denotes convergence with probability one (strong consistency).

Let us now consider the special case $r = \infty$. Write $N^+_n = N^+_{n,\infty}$, $\mathcal{B}^+_n = B^+_{n,\infty}$, $\psi^+ = \psi^+_\infty$, etc., so that, in particular, $N^+_n = \#\{\mathcal{B}^+_n\}$, $N^-_n = \#\{\mathcal{B}^-_n\}$, where

$$\mathcal{B}^+_n = \{b_{ij} \in \mathcal{B}_n : b_{ij} > \beta\}, \quad \mathcal{B}^-_n = \{b_{ij} \in \mathcal{B}_n : b_{ij} < \beta\}.$$

The two U-statistics are

$$\bar{N}^+_n = \frac{2}{n(n-1)}N^+_n = \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\mathbf{1}[(\epsilon_i - \epsilon_j)(x_i - x_j) > 0] \quad \text{and}$$
$$\bar{N}^-_n = \frac{2}{n(n-1)}N^-_n = \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}\mathbf{1}[(\epsilon_i - \epsilon_j)(x_i - x_j) < 0]$$

with kernels $\psi^+(Z_i, Z_j) = \mathbf{1}[(\epsilon_i - \epsilon_j)(x_i - x_j) > 0]$ and $\psi^-(Z_i, Z_j) = \mathbf{1}[(\epsilon_i - \epsilon_j)(x_i - x_j) < 0]$, respectively. Since $q = q_{n,\infty} = P(\epsilon_1 > \epsilon_2)$, (2.2) simplifies to

$$E\left\{\bar{N}^-_n\right\} = E\left\{\bar{N}^+_n\right\} = q\bar{a}_n = P(\epsilon_1 > \epsilon_2)\bar{a}_n. \tag{2.4}$$

Apparently $0 \leq q \leq 1/2$. In fact, we have a stronger result for a discontinuous cdf F.

Remark 1 *If F is discontinuous, then* $0 \leq q < 1/2$.

Indeed, if F is discontinuous at t_0, i.e., $F(t_0) - F(t_0-) > 0$, then

$$P(\epsilon_1 = \epsilon_2) = \int P(\epsilon_1 = t)\,dF(t) = \int (F(t) - F(t-))\,dF(t)$$
$$\geq (F(t_0) - F(t_0-))^2 > 0,$$

hence $q = P(\epsilon_1 > \epsilon_2) = (1/2)(1 - P(\epsilon_1 = \epsilon_2)) < 1/2$, proving Remark 1.

As an illustration, let us see an example. Suppose that ϵ_1 and ϵ_2 are independent Bernoulli trials with probability of success p. Then

$$P(\epsilon_1 = \epsilon_2) = P(\epsilon_1 = \epsilon_2 = 0) + P(\epsilon_1 = \epsilon_2 = 1) = (1-p)^2 + p^2 > 0.$$

The theorem below gives an interesting result when F is discontinuous, which implies the strong consistency. Let Ω_o be the set of all $\omega \in \Omega$ at which $\tilde{\beta}_n(\omega) \neq \beta$ happens only finitely many times, i.e., $\Omega_o = \{\omega : \tilde{\beta}_n(\omega) = \beta \text{ for } n > n_\omega\}$.

Theorem 1 *Suppose that F is discontinuous. If the nonrandom covariates x_1, x_2, ..., x_n satisfy (2.3), then $P\{\Omega_o\} = 1$. Consequently, the Theil-Sen estimator is strongly consistent, i.e., $\tilde{\beta}_n \to \beta$ a.s.*

Proof. Denote

$$\Omega_o' = \left\{\omega : \frac{N_n^+(\omega)}{a_n} \to q\ ,\ \frac{N_n^-(\omega)}{a_n} \to q \text{ and } a_n \to \infty\right\}.$$

Because of (2.3), we have $P(\Omega_o') = 1$ by Lemma 1. Fix $\omega \in \Omega_0'$ and set $q_0 = (\frac{1}{2} + q)/2$. Since F is discontinuous, we have $q < q_0 < 1/2$ by Remark 1. Thus, there exists an integer n_ω such that $N_n^+(\omega)/a_n < q_o$ and $N_n^-(\omega)/a_n < q_o$ whenever $n \geq n_\omega$, which implies that the median of $\mathcal{B}_n$ is inside the difference $\mathcal{B}_n \setminus (\mathcal{B}_n^- \cup \mathcal{B}_n^+)$. Otherwise, either $N_n^+(\omega)/a_n$ or $N_n^+(\omega)/a_n$ should be not less than 1/2. Since all random variables in $\mathcal{B}_n \setminus (\mathcal{B}_n^- \cup \mathcal{B}_n^+)$ are equal to β, it follows that the median $\tilde{\beta}_n$ of $\mathcal{B}_n$ satisfies $\tilde{\beta}_n(\omega) = \beta$ for $\omega \in \Omega_o'$. But $\Omega_o' \subset \Omega_o$ and $\mathrm{P}(\Omega_o') = 1$; hence, $\mathrm{P}(\Omega_o) = 1$. This completes the proof.

From the theorem we see surprisingly that the Theil-Sen estimator $\tilde{\beta}_n$ equals β for sufficiently large n almost surely when the cdf F is discontinuous. An immediate consequence of this is that $\tilde{\beta}_n$ is super-efficient as discussed in Case I of Section 3 below. Our simulation study also confirms this finding.

We now consider continuous cdf F. The next lemma is needed and its proof is given in appendix.

Lemma 2 *Suppose that F is continuous. Then the cdf F_2 of $\epsilon_1 - \epsilon_2$ satisfies*

$$F_2(t) = E[F(\epsilon + t)] = 1 - E[F(\epsilon - t)], \quad t \in \mathbb{R}, \tag{2.5}$$

hence $F_2(0) = 1/2$. Furthermore,

$$F_2(t) \neq 1/2, \ \text{for } t \neq 0. \tag{2.6}$$

Theorem 2 *Suppose that F is continuous. If the nonrandom covariates $x_1, x_2, \ldots$*

satisfy (2.3) and

$$\rho \equiv \lim_{n\to\infty} \inf \{|x_i - x_j| : x_i \neq x_j, 1 \leq i < j \leq n\} > 0, \tag{2.7}$$

then the Theil-Sen estimator is strongly consistent, i.e., $\tilde{\beta}_n \to \beta$ *a.s.*

Proof. Fix $r = 1, 2, \ldots$. It follows from Lemma 1 that

$$\lim_{n\to\infty} \left(N^+_{n,r}/a_n - q_{n,r}\right) = 0, \quad \lim_{n\to\infty} \left(N^-_{n,r}/a_n - q_{n,r}\right) = 0, \quad \text{a.s.}, \tag{2.8}$$

where $q_{n,r}$ is given in (2.1). Let $\Omega'_o = \bigcap_{r\geq 1} \Omega'_r$, where

$$\Omega'_r = \left\{\omega : N^+_{n,r}(\omega)/a_n - q_{n,r} \to 0, \;\; N^-_{n,r}(\omega)/a_n - q_{n,r} \to 0, \;\; a_n \to \infty\right\}.$$

Then by (2.8) we have $P(\Omega'_r) = 1$, $r = 1, 2, \ldots$, implying the complement satisfies $P(\Omega'^c_r) = 0$. Therefore,

$$P(\Omega'^c_o) = P\Big(\bigcup_{r\geq 1} \Omega'^c_r\Big) \leq \sum_{r\geq 1} P(\Omega'^c_r) = 0.$$

This proves $P(\Omega'_o) = 1$.

We observe that $q_{n,r} \leq 1 - F_2(\rho_n/r)$ holds for all r, n, where ρ_n is defined like ρ in (2.7) but without limit, i.e.,

$$\rho_n = \inf \{|x_i - x_j| : x_i \neq x_j, 1 \leq i \neq j \leq n\}.$$

Then we have

$$\limsup_{n\to\infty} q_{n,r} \leq 1 - \liminf_{n\to\infty} F_2(\rho_n/r) = 1 - F_2(\liminf_{n\to\infty} \rho_n/r) = 1 - F_2(\rho/r).$$

Since $\rho > 0$, it follows from Lemma 2 that $\limsup_{n\to\infty} q_{n,k} < 1/2$.

Analogous to the proof of Theorem 1, one may show that for each fixed r and each $\omega \in \Omega'_r$, there exists $n_{r,\omega}$ such that the median of $\mathcal{B}_n$ falls in the difference $\mathcal{B}_n \backslash (\mathcal{B}^-_{n,r} \cup \mathcal{B}^+_{n,r})$ for $n > n_{r,\omega}$. In other words, $|\tilde{\beta}_n(\omega) - \beta| \leq 1/r$. Letting $r \to \infty$ yields the desired strong consistency and the proof is complete.

3 Asymptotic Distribution

In this section, we study the asymptotic distribution of the Theil-Sen estimator for both discontinuous and continuous error cdf F. For discontinuous cdf F, we show that the Theil-Sen estimator is super-efficient; for continuous cdf F, we establish a general theorem on the asymptotic distribution. We provide two sets of sufficient conditions that satisfy the general theorem and permit explicit formulas of the scaling constants as well. We also present a sufficient condition for non-normal asymptotic distribution and an example of the error distribution F when that condition holds. We end this section with a small simulation study.

Case I Suppose that F is discontinuous. From the proof (at the end) of Theorem 1 we have

$$n^{\nu}(\tilde{\beta}_n - \beta) \to 0, \qquad \text{a.s.}$$

for every $\nu \in [0, \infty)$. This implies that $\tilde{\beta}_n$ is super-efficient in the case of discontinuous cdf F. A well-known super-efficient estimator was given by J. L. Hodges. For more details, see page 134 of Ferguson (1996). The simulation results at the end of this section confirm this asymptotic behavior.

Case II Now suppose that F is continuous. Denote

$$c_i = \sum_{j=1}^{n} \mathbf{1}[x_j > x_i] - \mathbf{1}[x_j < x_i], \quad C_n^2 = \sum_{i=1}^{n} c_i^2.$$

We now state the general theorem on the asymptotic distribution of the Theil-Sen estimator and defer its proof to the next section.

Theorem 3 *Suppose that* F *is continuous. If* $\{k_n\}$ *is such a sequence that* $k_n \to \infty$, *that* $\max_{1 \le i,j \le n} |x_i - x_j|/k_n \to 0$, *that* $\liminf_{n\to\infty} C_n/n^{3/2} > 0$, *and that*

$$\frac{1}{C_n} \sum_{1 \le i < j \le n} (1 - 2F_2(t|x_i - x_j|/k_n)) \to m(t), \quad t \in \mathbb{R}, \tag{3.1}$$

then

$$\lim_{n\to\infty} P\left\{k_n(\tilde{\beta}_n - \beta) \le t\right\} = \Phi(-\sqrt{3}m(t)), \quad t \in \mathbb{R}, \tag{3.2}$$

where Φ *is the cdf of the standard normal distribution* $\mathcal{N}(0,1)$.

Remark 2 *The asymptotic distribution in Theorem 3 is a composite of the standard normal distribution* $z \mapsto \Phi(z), z \in \mathbb{R}$ *and the function* $t \mapsto -\sqrt{3}m(t), t \in \mathbb{R}$. *Hence, it is not normal in general. It is normal if and only if* $m(t) = At, t \in \mathbb{R}$ *for some constant* $A \neq 0$.

We now turn to seek sufficient conditions that enable us to find explicit formulas for the scaling sequence $\{k_n\}$ in Theorem 3. To this end, the continuity of function $F_2(t)$ at 0 is not sufficient. We need $F_2(t) - 1/2 = F_2(t) - F_2(0)$ to tend to 0 fast enough. One such sufficient condition is as follows: There exist constants $\gamma > 0$ and $\alpha > 0$ such that

$$F_2(t) - 1/2 = \gamma \text{sign}(t)|t|^\alpha + o(|t|^\alpha), \quad t \in \Delta, \tag{3.3}$$

for some neighborhood Δ of the origin, where sign(t) is the sign function of t such that sign(t) $= 1$ if $t > 0$ and sign(t) $= -1$ if $t \leq 0$. The condition (3.3) appears in Smirnov (1952).

This condition is mild and satisfied by most of the common distributions. For example, Cauchy distribution function F with density $f(t) = 1/\pi(1 + t^2)$ satisfies (3.3) with $\gamma = 1/2\pi$ and $\alpha = 1$. We summarize some properties of (3.3) in Remark 3 below.

Remark 3 *(i) Apparently, condition (3.3) implies the continuity of* F *at zero; otherwise, for* $t > 0$, $F_2(t) - 1/2 \geq \frac{1}{2}P(\epsilon_1 = \epsilon_2) > 0$.
(ii) If F_2 *satisfies (3.3), then it is Hölder continuous with index* α*; if further* $0 < \alpha < 1$, *then* F_2 *is not differentiable at* 0.
(iii) F_2 *has a derivative* $F_2'(0)$ *at 0 if and only if (3.3) holds with* $\alpha = 1$ *for some constant* $\gamma > 0$*; thereby* $\gamma = F_2'(0)$.
(iv) If F *has a square-integrable density, then* $\gamma = F_2'(0) = B(F) = \int f^2(t)\, dt$.

The power distribution function with power less than 1/2 satisfies condition (3.3) with $0 < \alpha < 1$; see Example 1 below.

For $\alpha > 0$, let

$$d_i^\alpha = \sum_{j=1}^n |x_i - x_j|^\alpha, \quad D_n^\alpha = \sum_{i=1}^n d_i^\alpha, \quad b_n^\alpha = D_n^\alpha / C_n, \quad b_n = b_n^1, \quad D_n = D_n^1.$$

The first set of sufficient conditions that permits explicit formulas of the scaling sequence $\{k_n\}$ in Theorem 3 is stated in Theorem 4.

Theorem 4 *Suppose that F is continuous. Assume that there exist constants $\gamma > 0$, $\alpha > 0$ such that (3.3) holds. If $b_n^\alpha \to \infty$, $\max_{1\le i,j\le n} |x_i - x_j|^\alpha / b_n^\alpha \to 0$ and $\liminf_{n\to\infty} C_n/n^{3/2} > 0$, then*

$$(b_n^\alpha)^{1/\alpha}(\tilde{\beta}_n - \beta) \Longrightarrow \xi_{\alpha,\gamma},$$

where $P(\xi_{\alpha,\gamma} < t) = \Phi(\sqrt{3}\gamma \text{sign}(t)|t|^\alpha)$, $t \in \mathbb{R}$ and $\Longrightarrow$ stands for convergence in distribution.

Proof. The result follows from an application of Theorem 3 with $k_n = (b_n^\alpha)^{1/\alpha}$ and $m(t) = -\gamma \text{sign}(t)|t|^\alpha$.

For a related result, see page 150 of Koenker (2005) for non-normal asymptotic distributions of quantiles. An immediate corollary of Theorem 4 follows.

Corollary 1 *Suppose that the assumptions of Theorem 4 hold and that F_2 has a positive derivative d at 0. Then*

$$(D_n/C_n)(\tilde{\beta}_n - \beta) \Longrightarrow \mathcal{N}(0,\ 1/(3d^2)).$$

Now we consider the second set of sufficient conditions that ensures an explicit formula for the scaling sequence $\{k_n\}$ in Theorem 3. Suppose that F is absolutely continuous with a density function f that is square-integrable with respect to the Lebesgue measure, i.e.,

$$B(F) = \int f^2(t)\, dt < \infty. \tag{3.4}$$

Under this condition we can calculate the limit $m(t) = -tB(F)$ and have the explicit formula $k_n = D_n/C_n$. Thus we have shown the following theorem, which can also be viewed as a generalization of Corollary 1.

Theorem 5 *Suppose that F is absolutely continuous with a density function f such that (3.4) is true. If $\liminf C_n/n^{3/2} > 0$, then*

$$\frac{D_n}{C_n}(\tilde{\beta}_n - \beta) \Longrightarrow \mathcal{N}\left(0,\ \frac{1}{3B^2(F)}\right).$$

This theorem corresponds to Theorem 6.2 of Sen (1968). Sen (page 1387) claimed that the asymptotic normality of the Theil-Sen estimator follows from an application of Theorem 7.1 of Hoeffding (1948), but his setting does not satisfy the condition that the random variables have an identical distribution as required in Theorem 7.1.

Suppose now that the covariates $x_1, \ldots, x_n$ are distinct. Then we can calculate C_n^2 as follows: Since $c_i = \sum_{j=1}^n \mathbf{1}[x_j > x_i] - \mathbf{1}[x_j < x_i] = (n-i) - (i-1)$, it follows that

$$C_n^2 = \sum_{i=1}^n c_i^2 = \sum_{i=1}^n [(n-i) - (i-1)]^2 = n(n^2-1)/3.$$

Suppose now that there are ties among the covariates $x_1, \ldots, x_n$. Then C_n^2 shall be modified as $n(n^2-1)/3$, minus the number of repetitions caused by those tied x_i's that were counted as distinct values. Suppose that there are r_n distinct values $x'_1, \ldots, x'_{r_n}$ among $x_1, \ldots, x_n$. We partition the covariate set $\{x_1, \ldots, x_n\}$ into r_n subsets S_k, with each S_k containing all x_i that equal x'_k. Suppose that S_k has u_k elements. For $x_i \in S_k$ there are three types of repetitions resulting from S_k in calculating c_i: the first type of repetitions come from $\sum_{j=1}^n \mathbf{1}[x_j > x_i]$, the second from $\sum_{j=1}^n \mathbf{1}[x_j < x_i]$ and the third from $\sum_{x_j \in S_k} \mathbf{1}[x_j < x_i] - \mathbf{1}[x_j < x_i]$. The first and second types cancel out in c_i and the third type has $u_k(u_k^2-1)/3$ elements (which is calculated from the above when $n = u_k$). Therefore, we have

$$C_n^2 = \sum_{i=1}^n c_i^2 = \frac{1}{3}\left(n(n^2-1) - \sum_{k=1}^{r_n} u_k(u_k^2-1)\right).$$

An algebraic proof of this equality can also be derived from the identity (5.7) given in the appendix.

Next we have a special case as a corollary.

Corollary 2 *Suppose that F is absolutely continuous with a square-integrable density function f. If the covariates $x_1, ..., x_n$ are distinct, then*

$$\frac{D_n}{n^{3/2}}(\tilde{\beta}_n - \beta) \Longrightarrow \mathcal{N}\left(0, \frac{1}{9B^2(F)}\right).$$

From Theorem 4, we see that if $\alpha \neq 1$, then the asymptotic distribution of the Theil-Sen estimator is not normal. An important question to ask is whether a distribution F exists for $0 < \alpha < 1$. The following example answers this question positively.

Example 1 Consider the power distribution function $F(t) = t^\theta, 0 \leq t \leq 1$, with $0 < \theta < 1/2$. From (2.5) we have $F_2(t) = \int F(s+t)\, dF(s) = \int_0^{1-t}(s+t)^\theta\, dF(s)$ and $F_2(t) - F_2(0) = F_2(t) - 1/2 = \int_0^{1-t}(s+t)^\theta\, ds^\theta - \int_0^1 s^\theta\, ds^\theta$. As a result, it is easy to see that F_2 is not differentiable at 0 for $0 < \theta < 1/2$. Indeed, F_2 is differentiable at 0 if and only if $\theta > 1/2$. But condition (3.3) is satisfied with $\alpha = 2\theta$ for $0 < \theta < 1/2$ and $\gamma = \int_0^\infty [(1 + x^{1/\theta})^\theta - x]\, dx$ as is shown below. For $0 < t < 1$,

$$\begin{aligned}
\frac{F_2(t) - 1/2}{t^{2\theta}} &= \int_0^{1-t} \frac{(s+t)^\theta - s^\theta}{t^{2\theta}}\, ds^\theta - \int_{1-t}^1 \frac{s^\theta}{t^{2\theta}}\, ds^\theta \\
&= \int_0^{(1-t)^\theta/t^\theta} [(1 + u^{1/\theta})^\theta - u]\, du + o(1) \\
&= \int_0^\infty [(1 + u^{1/\theta})^\theta - u]\, du, \quad t \to 0.
\end{aligned}$$

The last improper integral is convergent because

$$(1 + u^{1/\theta})^\theta - u = u[(1 + u^{-1/\theta})^\theta - 1] = O(u^{1-1/\theta}), \quad u \to \infty,$$

while $\int_1^\infty u^{1-1/\theta}\, du < \infty$ as $0 < \theta < 1/2$. Combining the above yields the desired result. It is interesting to point out that $\theta = 1/2$ is a breaking point at which neither F_2 is differentiable at 0 nor condition (3.3) is satisfied.

A small simulation study. We conduct two simulations to confirm (1) the super-efficiency and (2) non-normality of the asymptotic distribution.

(1) In order to investigate how large the sample size n needs to be such that $\tilde{\beta}_n = \beta$ in the regression Model (1.1) with discontinuous errors, we conduct the following simulation. For sample size n, we first generate the covariates $x_1, \ldots, x_n$ from some distribution and treat them as nonrandom fixed numbers. For this fixed set of covariates $\{x_1, \ldots, x_n\}$, we then generate $N = 500$ samples of random errors $\epsilon_1, \ldots, \epsilon_n$ from some discontinuous distribution. For each sample, we calculate the responses Y_i with a chosen β value according to Model (1.1) and the Theil-Sen estimator $\tilde{\beta}_n$. Based on the calculation from

the 500 samples, we compute the proportion of these samples that $\tilde{\beta}_n = \beta$. We use various distributions for the covariates and random errors. The results are reported in Table 1. In the upper part of the table, the covariates x_i are generated from binomial (Bin) distribution with size 100 and probability 0.75 of success, while the random errors are generated from uniform distribution on the set $\{-1, 1\}$ (± 1), Poisson distribution with unit mean (Pios) and binomial distribution with size 100 and probability 0.75 of success, respectively. In the table's lower left panel, the covariates are generated from the standard normal distribution ($\mathcal{N}$) and the errors from the same aforementioned three error distributions, whereas in the lower right panel, the errors are generated from Poisson distribution with unit mean and the covariates are generated from binomial distribution with size 100 and probability 0.75 of success, standard normal distribution and negative binomial (NB) distribution with size 5 and probability 0.4 of success, respectively. From the table we see that the value of n, at which $\tilde{\beta}_n = \beta$, relies tremendously on the distributions of the covariates and random errors. For example, in the upper part of the table, where the covariates x_i are generated from binomial distribution, the proportion reaches 1 at sample size $n = 50$ for ± 1 and Poisson random errors, while it only reaches 0.910 at $n = 400$ for binomial random errors. It is worthwhile to point out that these simulation results should be interpreted conditionally given the particular set of covariate values x_i.

(2) Similar to the above, we also generate 500 samples with sample size $n = 200$ according to Model (1.1), with covariates x_i generated from the standard normal distribution (treated as nonrandom) and random errors from the power distribution $F(t) = t^{1/4}$ in Example 1. The true value of β is 1. We calculate the responses for $\beta = 1$ and the estimator $\tilde{\beta}_n$ for each sample. Figure 1 shows the normality plot and the histogram of the simulated Theil-Sen estimator. The normality plot is S-shaped, indicating that the sampling distribution is extremely heavy-tailed. In addition, both the Shapiro and Kolmogorov-Smirnov tests of normality give 0 p-value, supporting the non-normal asymptotic distribution of the estimator.

Table 1

Proportions of $\tilde{\beta}_n = \beta$ with $N = 500$ replicates and different sample sizes.

x	Err	5	20	50	80	100	150	250	400
Bin	± 1	0.564	0.998	1.000	1.000	1.000	1.000	1.000	1.000
	Pois	0.222	0.932	1.000	1.000	1.000	1.000	1.000	1.000
	Bin	0.034	0.270	0.472	0.580	0.664	0.724	0.862	0.910

x	Err	5	20	50	80	Err	x	5	20	50	80
$\mathcal{N}$	± 1	.484	.998	1.000	1.000	Pois	Bin	.290	.930	.996	1
	Pois	.252	.942	0.998	1.000		$\mathcal{N}$	.212	.918	.996	1
	Bin	.028	.312	0.514	0.592		NB	.268	.932	.994	1

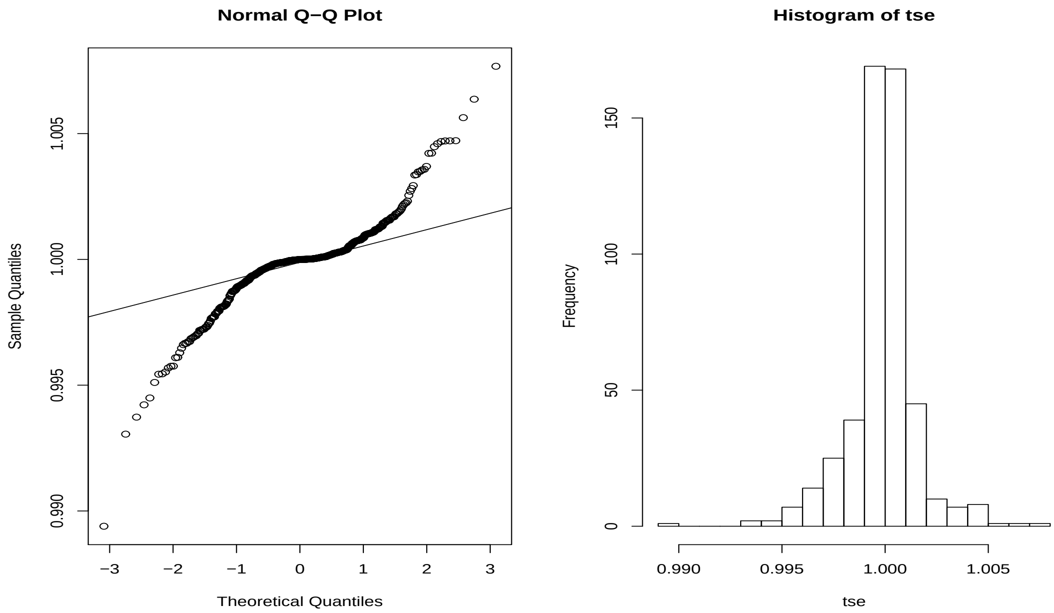

Fig. 1. Simulated Theil-Sen estimator $\tilde{\beta}_n$. Left: Normality plot. Right: Histogram. Both plots indicate the non-normality.

4 Proof of Theorem 3

In the proof of Theorems 1 and 2, we constructed several U-statistics to establish the strong consistency of the Theil-Sen estimator $\tilde{\beta}$. One significant feature of our U-statistics is that the kernels vary with the sample size n, which presents us some technical challenges. These U-statistics are different from the so called Kendall's tau, which was used by Sen (1968). Furthermore, when the covariates $x_1, \ldots, x_n$ are nonidentical nonrandom constants, neither our U-statistics nor the Kendall's tau are the usual U-statistics which do not depend on the sample size n. We remind the readers here the notation $Z_1 = (x_1, \epsilon_1), \ldots, Z_n = (x_n, \epsilon_n)$ are independent but not identically distributed random vectors. When establishing the asymptotic normality of the Theil-Sen estimator, Sen (1968) applied a theorem from Hoeffding (1948) that requires these Z_i's to be independent and identically distributed random vectors.

Based on $Z_1, \ldots, Z_n$, we define a U-statistic $V_n(t)$ by

$$V_n(t) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \psi(Z_i, Z_j; t), \quad t \in \mathbb{R}, \tag{4.1}$$

with kernel

$$\psi(Z_i, Z_j; t) = (\mathbf{1}[(\epsilon_i - \epsilon_j)/(x_i - x_j 2) > t] - 1/2)\, \mathbf{1}[x_i \neq x_j]. \tag{4.2}$$

The asymptotic behavior of $V_n(t_n)$ as $t_n \to 0$ is given below and the proof is defered to the last section. Schick and Wefelmeyer (2004) investigate the asymptotic behavior of U-statistics with kernel depending on the sample size, but with iid $Z_1, ..., Z_n$.

Before we prove Theorem 3, let us introduce two useful lemmas.

Lemma 3 *Suppose that F is continuous. Then for every sequence $\{t_n\}$ such that*

$$t_n \max_{1 \le i,j \le n} |x_i - x_j| \to 0, \tag{4.3}$$

we have

$$V_n(t_n) = \mu_n(t_n) + \frac{1}{n^2} \sum_{i=1}^{n} \Big(1 - 2F(\epsilon_i)\Big) c_i + o_p(n^{-1/2}),$$

where

$$\mu_n(t_n) = \frac{1}{n^2} \sum_{1 \le i < j \le n} (1 - 2F_2(|x_i - x_j| t_n)) \, .$$

Lemma 4 *(Wang 2005) Let $y_1, \ldots, y_N$ be real numbers and y_{med} be their median. If N is an odd integer, then*

$$y_{med} > t \Leftrightarrow \# \{i : \ y_i > t\} \geq N/2, \ t \in \mathbb{R};$$

If N is an even integer, then

$$\# \{i : \ y_i > t\} \geq (N+1)/2 \Rightarrow y_{med} > t \Rightarrow \# \{i : \ y_i > t\} \geq N/2, \ t \in \mathbb{R}.$$

We are now ready to show Theorem 3.

Proof of Theorem 3. Fix $t \in \mathbb{R}$ and let $t_n = tk_n^{-1}$. From Lemma 3, we have the stochastic expansion

$$\frac{\sqrt{3}n^2}{C_n} \left[V_n(t_n) - \mu_n(t_n)\right] = \sum_{i=1}^{n} \frac{\sqrt{3}c_i}{C_n}[1 - 2F(\epsilon_i)] + \frac{n^{3/2}}{C_n} o_p(1).$$

Since $F(\epsilon_i)$ is uniformly distributed on interval (0,1), it follows that $E\{F(\epsilon_i)\} = 1/2$ and $Var\{F(\epsilon_i)\} = 1/12$. In view of $\liminf_{n\to\infty} C_n/n^{3/2} > 0$, we have the asymptotic normality

$$\frac{\sqrt{3}n^2}{C_n} \left[V_n(t_n) - \mu_n(t_n)\right] \Longrightarrow \mathcal{N}(0, 1). \tag{4.4}$$

Here "$\Longrightarrow$" denotes convergence in distribution. Write $t_n = tk_n^{-1}$ and

$$P\left\{k_n(\tilde{\beta}_n - \beta) > t\right\} = P\left\{\tilde{\beta}_n - \beta > t_n\right\}.$$

Recall that $\tilde{\beta}_n = \text{med}(\mathcal{B}_n) = \text{med}(\{b_{ij} : x_i \neq x_j, 1 \leq i < j \leq n\})$ and

$$b_{ij} = (Y_i - Y_j)/(x_i - x_j) = \beta + (\epsilon_i - \epsilon_j)/(x_i - x_j) = \beta + e_{ij}, \quad x_i \neq x_j.$$

From Lemma 4, we have, with $N_n = n(n-1)/2$,

$$P\left\{k_n(\tilde{\beta}_n - \beta) > t\right\} \leq P\left\{\sum_{1\leq i<j\leq n} \mathbf{1}[e_{ij} > t_n]\mathbf{1}[x_i \neq x_j] \geq N_n/2\right\},$$

and

$$P\left\{k_n(\tilde{\beta}_n - \beta) > t\right\} \geq P\left\{\sum_{1\leq i<j\leq n} \mathbf{1}[e_{ij} > t_n]\mathbf{1}[x_i \neq x_j] \geq \frac{N_n + 1}{2}\right\}.$$

From (4.1) and (4.2),

$$\sum_{1\le i<j\le n} \mathbf{1}[e_{ij} > t_n]\mathbf{1}[x_i \ne x_j] - N_n/2 = \sum_{1\le i<j\le n} \Big(\mathbf{1}[e_{ij} > t_n] - 1/2\Big)\mathbf{1}[x_i \ne x_j]$$
$$= \sum_{1\le i<j\le n} \psi(Z_i, Z_j; t_n) = \frac{n(n-1)}{2} V_n(t_n).$$

Thus,

$$P\left\{k_n(\tilde{\beta}_n - \beta) > t\right\} \le P\left\{\frac{\sqrt{3}n^2}{C_n}\left[V_n(t_n) - \mu_n(t_n)\right] \ge -\frac{\sqrt{3}n^2}{C_n}\mu_n(t_n)\right\},$$

and

$$P\left\{k_n(\tilde{\beta}_n - \beta) > t\right\} \ge P\left(\frac{\sqrt{3}n^2}{C_n}\left[V_n(t_n) - \mu_n(t_n)\right] \ge -\frac{\sqrt{3}n^2}{C_n}\mu_n(t_n) + \frac{\sqrt{3}}{\sqrt{n}C_n}\right).$$

The last two equalities, (4.4) and (3.1) yield the desired (3.2) and the proof is complete.

5 Appendix

Here we provide the proofs of some lemmas that were used in the proofs of Theorems 1–4. Lemma 1 is a SLLN for the U-statistics used in the proofs of Theorems 1 and 2. Lemma 2 states an interesting property of the cdf F_2 of the difference of two iid random errors. Lemma 3 gives the asymptotic normality of the U-statistic $V_n(t_n)$ used in the proof of Theorem 4.

Proof of Lemma 1. We only show the first limit because the second is similar. Observing that $\bar{N}^+_{n,r}$ is a U-statistic with a bounded kernel $\psi^+_r(Z_1, Z_2)$, an application of the Hoeffding inequality for a U-statistic (inequality (5.7) of Hoeffding, 1963) yields

$$P\left(\bar{N}^+_{n,r} - E(\bar{N}^+_{n,r}) \ge t\right) \le \exp(-nt^2), \quad t > 0.$$

It follows that

$$P\left(\left|\bar{N}^+_{n,r}/\bar{a}_n - q_{n,r}\right| \ge t\right) \le 2\exp\left(-n\bar{a}_n^2 t^2\right), \quad t > 0. \tag{5.1}$$

In order to show the first limit using the Borel-Cantelli lemma, it suffices to show that for every $t > 0$,

$$\sum_{n=2}^{\infty} P\left(\left|\bar{N}_{n,r}^{+}/\bar{a}_n - q_{n,r}\right| \geq t\right) < \infty. \tag{5.2}$$

Since $n^{-1}\log n/\bar{a}_n^2 = o(1)$, it follows that for every arbitrarily fixed $t > 0$, there exists n_t such that $n\bar{a}_n^2 t^2 \geq 2\log n$ for $n > n_t$, therefore

$$\sum_{n=n_t}^{\infty} \exp\left(-n\bar{a}_n^2 t^2\right) \leq \sum_{n=n_t}^{\infty} \exp\left(-2\log n\right) = \sum_{n=n_t}^{\infty} \frac{1}{n^2} < \infty.$$

Hence, (5.1) implies (5.2) and the proof is complete.

Proof of Lemma 2. Inequality (2.6) is from Wang (2005) and (2.5) follows from

$$F_2(t) = P(\epsilon_1 - \epsilon_2 \leq t) = \int P(\epsilon_1 \leq t_2 + t)\, dF(t_2) = E[F(\epsilon + t)], \quad \text{and}$$

$$F_2(t) = P(\epsilon_1 - \epsilon_2 \leq t) = \int P(\epsilon_2 \geq t_1 - t)\, dF(t_1) = 1 - E[F(\epsilon - t)],$$

where $\epsilon, \epsilon_1, \epsilon_2$ are iid with common continuous cdf F. This completes the proof.

Proof of Lemma 3. We calculate

$$\begin{aligned}
\mu_n(t_n) = E(V_n(t_n)) &= \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} E(\psi(Z_i, Z_j; t_n)) \\
&= \frac{2}{n(n-1)} \sum_{i<j} E\left(\mathbf{1}\left[(\epsilon_i - \epsilon_j)/(x_i - x_j) > t_n\right] - 1/2\right) \mathbf{1}\left[x_i \neq x_j\right] \\
&= \frac{2}{n(n-1)} \sum_{i<j} \left[P\Big((\epsilon_i - \epsilon_j)/(x_i - x_j) > t_n\Big) - 1/2\right] \mathbf{1}\left[x_i \neq x_j\right] \\
&= \frac{2}{n(n-1)} \sum_{i<j} \left[1/2 - F_2(|x_i - x_j| t_n)\right] \mathbf{1}\left[x_i \neq x_j\right].
\end{aligned}$$

For $z = (x, \epsilon) \in \mathbb{R}^2$ and $t \in \mathbb{R}$, let

$$\psi_{ni}(z, t) = \frac{1}{n-1} \sum_{j\neq i} E\left(\psi(z, Z_j, t)\right), \quad V_{n1}(t) = \frac{1}{n} \sum_{i=1}^{n} \psi_{ni}(Z_i, t),$$

and

$$R_n(t) = \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} \psi(Z_i, Z_j, t) - 2\psi_{ni}(Z_i, t) + \mu_n(t).$$

Since ψ is a symmetric kernel bounded by 1, it follows that $R_n(t) = o_p(n^{-1/2})$. The Hoeffding decomposition of U-statistic $V_n(t_n)$ is

$$V_n(t_n) = \mu_n(t_n) + 2(V_{n1}(t_n) - \mu_n(t_n)) + o_p(n^{-1/2}). \tag{5.3}$$

Next we show that

$$D_n(t_n) \equiv V_{n1}(t_n) - \mu_n(t_n) - V_{n1}(0) = o_p(n^{-1/2}). \tag{5.4}$$

By straightforward computation we find

$$\psi_{ni}(z,t) = \frac{1}{n-1}\sum_{j\neq i}\Big(1/2 - F(\epsilon + (x_j - x)t)\Big)\Big(\mathbf{1}[x_j > x] - \mathbf{1}[x_j < x]\Big).$$

In particular,

$$\begin{aligned}\psi_{ni}(Z_i,0) &= \frac{1}{n-1}\sum_{j\neq i}\Big(1/2 - F(\epsilon_i)\Big)\Big(\mathbf{1}[x_j > x_i] - \mathbf{1}[x_j < x_i]\Big)\\ &= \Big(1/2 - F(\epsilon_i)\Big)\frac{c_i}{n-1},\end{aligned}$$

and

$$V_{n1}(0) = \frac{1}{n}\sum_{i=1}^{n}\psi_{ni}(Z_i,0) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\Big(1/2 - F(\epsilon_i)\Big)c_i. \tag{5.5}$$

Next we show that

$$nE\left[D_n^2(t_n)\right] \le K\Big(F_2(|t_n|\max_{i,j}|x_i - x_j|) - F_2(-|t_n|\max_{i,j}|x_i - x_j|)\Big) + K/n \tag{5.6}$$

for some constant K. This inequality, Lebesgue dominated convergence theorem, Markov inequality, the continuity of F and hence of F_2, and (4.3) together imply (5.4). Furthermore, (5.4) together with (5.5) and (5.3), in turn, yields the desired result of the lemma.

It is left to show (5.6). It is easy to check that

$$\mu_n(t_n) = \frac{2}{n(n-1)}\sum_{i\neq j}\left[1/2 - F_2(x_{ji}t_n)\right]\mathbf{1}\left[x_j > x_i\right],$$

and

$$V_{n1}(t_n) = \frac{2}{n(n-1)}\sum_{i\neq j}(1/2)[F(\epsilon_j - x_{ji}t_n) - F(\epsilon_j + x_{ji}t_n)]\mathbf{1}[x_j > x_i],$$

hence, with $x_{ji} = x_j - x_i$,

$$V_{n1}(0) = \frac{2}{n(n-1)} \sum_{i \neq j} (1/2)[F(\epsilon_j) - F(\epsilon_j)] \mathbf{1}[x_j > x_i].$$

Using these and (2.5), we write $\binom{n}{2} D_n(t_n) = A_n - B_n$ with $A_n = \sum_{i=1}^n A_{ni} = \sum_{i=1}^n \sum_{j \neq i} A_{nij}$ and $B_n = \sum_{i=1}^n B_{ni} = \sum_{i=1}^n \sum_{j \neq i} B_{nij}$, where

$$A_{nij} = (1/2)\Big(F(\epsilon_j - x_{ji}t_n) - F(\epsilon_j) - E[F(\epsilon_j - x_{ji}t_n) - F(\epsilon_j)]\Big)\mathbf{1}[x_j > x_i],$$

$$B_{nij} = (1/2)\Big(F(\epsilon_j + x_{ji}t_n) - F(\epsilon_j) - E[F(\epsilon_j + x_{ji}t_n) - F(\epsilon_j)]\Big)\mathbf{1}[x_j > x_i].$$

Then

$$\begin{aligned}
&E\,(A_{nij}) = E\,(B_{nij}) = 0, \quad Cov\,(A_{nij_1}, A_{nij_2}) = 0, \quad j_1 \neq j_2, \\
&Var\,(A_{nij}) \leq KE\,(F(\epsilon - x_{ji}t_n) - F(\epsilon))^2 \\
&\qquad\qquad \leq K\,(F_2(|x_{ji}t_n|) - F_2(-|x_{ji}t_n|)) \equiv K\Delta_{ij}, \\
&Var\,(B_{nij}) \leq K\Delta_{ij}, \quad E\,(A_{ni_1j_1}A_{ni_2j_2}) = 0, \quad j_1 \neq j_2, \\
&E\,(A_{ni_1j}A_{ni_2j}) = Cov\Big(F(\epsilon_j - x_{ji_1}t_n) - F(\epsilon_j),\ F(\epsilon_j - x_{ji_2}t_n) - F(\epsilon_j)\Big) \\
&\qquad\qquad\qquad \leq 2\Delta_{i_1j}^{1/2}\Delta_{i_2j}^{1/2},
\end{aligned}$$

thus,

$$E\,(A_{ni_1}A_{ni_2}) = \sum_{j_1 \neq i_1} \sum_{j_2 \neq i_2} E\,(A_{ni_1j_1}A_{ni_2j_2}) \leq Kn \max_{i,j}(\Delta_{ij}),$$

where K is a general constant that may denote different values in different places. Hence,

$$E\left(A_{ni}^2\right) = \sum_{j \neq i} Var\,(A_{nij}) + \sum_{j_1 \neq j_2 \neq i} Cov\,(A_{nij_1}, A_{nij_2}) \leq Kn,\ E\left(B_{ni}^2\right) \leq Kn.$$

Therefore,

$$E\left(A_n^2\right) = \sum_i E\left(A_{ni}^2\right) + \sum_{i_1 \neq i_2} E\,(A_{ni_1}A_{ni_2}) \leq Kn^2 + Kn^3 \max_{i,j}(\Delta_{ij}).$$

Similarly, $E\,(B_n^2) \leq Kn^2 + Kn^3 \max_{i,j}(\Delta_{ij})$. Combining this with the above inequality yields the desired (5.6) and completes the proof.

Identity: For real numbers $u_1, u_2, ...$ and positive integer K, we have

$$\frac{1}{3}\left[\left(\sum_{i=1}^{K} u_i\right)^3 - \sum_{i=1}^{K} u_i^3\right] = \sum_{i=1}^{K} u_i \left(\sum_{j=i+1}^{K} u_j - \sum_{i=1}^{i-1} u_j\right)^2. \tag{5.7}$$

Proof. We apply induction on K. For $K = 1$ the identity is obvious. Suppose that the identity (5.7) holds for $K = k$. For $K = k + 1$, the left hand side of (5.7) is

$$\begin{aligned}
&\frac{1}{3}\left[\left(\sum_{i=1}^{k+1} u_i\right)^3 - \sum_{i=1}^{k+1} u_i^3\right] \\
&= \frac{1}{3}\left[\left(\sum_{i=1}^{k} u_i\right)^3 + 3u_{k+1}\left(\sum_{i=1}^{k} u_i\right)^2 + 3u_{k+1}^2 \sum_{i=1}^{k} u_i + u_{k+1}^3 - \sum_{i=1}^{k+1} u_i^3\right] \\
&= \frac{1}{3}\left[\left(\sum_{i=1}^{k} u_i\right)^3 - \sum_{i=1}^{k} u_i^3\right] + u_{k+1}\left(\sum_{i=1}^{k} u_i\right)^2 + u_{k+1}^2 \sum_{i=1}^{k} u_i,
\end{aligned}$$

and the right hand side is

$$\begin{aligned}
&\sum_{i=1}^{k+1} u_i \left(\sum_{j=i+1}^{k+1} u_j - \sum_{j=1}^{i-1} u_j\right)^2 \\
&= \sum_{i=1}^{k} u_i \left(u_{k+1} + \sum_{j=i+1}^{k} u_j - \sum_{j=1}^{i-1} u_j\right)^2 + u_{k+1}\left(\sum_{i=1}^{k} u_i\right)^2 \\
&= u_{k+1}^2 \sum_{i=1}^{k} u_i + 2u_{k+1} \sum_{i=1}^{k} u_i \left(\sum_{j=i+1}^{k} u_j - \sum_{j=1}^{i-1} u_j\right) \\
&\quad + \sum_{i=1}^{k} u_i \left(\sum_{j=i+1}^{k} u_j - \sum_{j=1}^{i-1} u_j\right)^2 + u_{k+1}\left(\sum_{i=1}^{k} u_i\right)^2 \\
&= \sum_{i=1}^{k} u_i \left(\sum_{j=i+1}^{k} u_j - \sum_{j=1}^{i-1} u_j\right)^2 + u_{k+1}\left(\sum_{i=1}^{k} u_i\right)^2 + u_{k+1}^2 \sum_{i=1}^{k} u_i.
\end{aligned}$$

By induction, the identity is correct.

Acknowledgments This research is supported in part by grant T32MH-014235 from the National Institute of Health. The authors would like to express their sincere thanks to an Associated Editor and the referees for helpful suggestions and comments that improved the presentations.

References

Arcones, M. A., Chen, Z., Giné, E., 1994. Estimators related to U-Processes with applications to multivariate medians: asymptotic normality. Annals of Statistics 22, 1460–1477.

Akritas, M. G., Murphy, S. A., LaValley, M. P., 1995. The Theil-Sen estimator with doubly censored data and applications to astronomy. Journal of the American Statistical Association 90, 170–177.

Bose, A., 1998. Bahadur representation of M_m estimates. Annals of Statistics 26, 771–777.

Dietz, E. J., 1987. A comparison of robust estimators in simple linear regression. Communications in Statistics, Part B—Simulation and Computation 16, 1209–1227.

Ferguson, T. S., 1996. A course in large sample theory. First edition, Chapman & Hall, New York.

Fernandes, R., Leblanc, S. G., 2005. Parametric (modified least squares) and non-parametric (Theil-Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. Remote Sensing of Environment 95, 303–316.

Hoeffding, W., 1948. A class of statistics with asymptotically normal distribution. Annals of Mathematical Statistics 19, 293–325.

Hoeffding, W., 1963. Probability inequalities for sums of bounded variables. Journal of the American Statistical Association 50, 13–30.

Hollander, M., Wolfe, D. A., 1973. Nonparametric statistical methods. First edition, John Wiley & Sons, New York, 205–206.

Hollander, M., Wolfe, D. A., 1999. Nonparametric statistical methods. Second edition, John Wiley & Sons, New York, 421–423.

Jones, M. P., 1997. A class of semi-parametric regressions for the accelerated failure time model. Biometrika 84, 73–84.

Koenker, R., 2005. Quantile regression. Cambridge University Press, Cambridge.

Mount, D. M., Netanyahu, N. S., 2001. Efficient randomized algorithms for robust estimation of circular arcs and aligned ellipses. Computational geometry: theory and applications 19, 1–33.

Oja, H., 1983. Descriptive statistics for multivariate distributions. Statistics and Probability Letters 1, 327–332.

Oja, H., Niinimaa, A., 1984. On robust estimation of regression coefficients. Research Report, Department of Applied Mathematics and Statistics, University of Oulu, Finland.

Rousseeuw, P. J., Leroy, A. M., 2003. Robust regression and outlier detection. John Wiley & Sons, New York.

Schick, A., Wefelmeyer, W., 2004. Root n consistent and optimal density estimators for moving average processes. Scandinavian Journal of Statistics 31, 63–78.

Sen, P. K., 1968. Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association 63, 1379–1389.

Sprent, P., 1993. Applied nonparametric statistical methods. Second Edition, CRC Press, New York.

Smirnov, N. V., 1952. Limit distributions for the terms of a variational series. American Mathematical Society Translations 67, 82–143.

Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis. Nederlandse Akademie Wetenchappen Series A. 53, 386–392.

Wang, X. Q., 2005. Asymptotics of the Theil-Sen estimator in simple linear regression models with a random covariate. Nonparametric Statistics 17, 107–120.

Wilcox, R. R., 1998. Simulations on the Theil-Sen regression estimator with right-censored data. Statistics and Probability Letters 39, 43–47.

Zhou, W., and Serfling, R., 2007. A Bahadur-Kiefer representation for multivariate spatial U-quantiles. Preprint.