ON A CONSISTENT RANK ESTIMATE
IN A LINEAR STRUCTURAL MODEL

Silvelyn Zwanzig

ABSTRACT. The structural linear model is considered that is an errors-in-variables model where the unobserved variables are i.i.d. In this model we can find linear transformations depending on the parameter, such that the transformed observations using the true parameter are uncorrelated. Then a parameter estimator is defined as a zero point of a consistent correlation estimator. A rank estimation is proposed as a zero point of Kendall's correlation measure and its consistency is shown. While the Pearson estimate of the covariance delivers the total least squares estimate.

1. Introduction

Rank methods are mostly used for constructing tests. They have the advantages that the null distribution of rank test statistics is independent of the underlying distribution. Rank estimation are unfortunately not applied in the same range, even though their theoretical properties are well known, see [5]–[7] and [9].

Furthermore the properties of rank methods in measurement errors models are now of interest, see [8]. In [12], where S e n has introduced the Theil–Sen estimate, he also discussed the robust behavior of his estimate under the special errors-in-variables set up, when the errors of variables are smaller than the double distances between two neighbored unobserved variables. In usual errors-in-variables models this restriction is not fulfilled. In [13] S e n and S a l e h have shown that the Theil–Sen estimator has the same bias as the naive least squares estimator in errors-in-variables models.

In Subsection 2.1 we remind the argumentation of S e n in [12]. His main idea is that Kendall's tau between the residuals and the x-variables should be zero

for an good estimate. In Subsection 2.2 we generalize this approach to structural errors-in-variables models. In an errors-in-variable set up the observable x-variables and the naive residuals given the true parameter are not uncorrelated. We use an idea of T h a n P e, [14] and consider instead of the observable x-variables a statistic which is orthogonal to the naive residuals. In Subsection 3.1 the rank estimate is introduced as the parameter value, where Kendall's tau between these statistics is zero. The consistency of the related rank estimate is shown. In Subsection 3.2 it is proved that the estimate based on Pearson covariance coincides with the total least squares estimate. These equivalence gives a further justification of the chosen approach, because under normal distribution the least squares estimator is the maximum likelihood estimate in the regression model and the total least squares estimate is the maximum likelihood estimate in the related errors-in-variables model. In Section 4 a small simulation is carried out as an evidence that the calculation of rank estimates is not any more problematic. The longer and technical proofs are given in the Appendix.

2. Models

In order to explain the main principle we consider first a multiple regression model with random design points and no errors in variables.

2.1. Multiple regression

Consider random vectors $z_i^T = \left(x_i^T, y_i \right)$, $i = 1, \ldots, n$, $x_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}$ fulfilling

$$y_i = \beta_0^T x_i + \varepsilon_i,$$

where x_i, ε_i are i.i.d. with $E\varepsilon_i = 0$, $E\varepsilon_i x_i = 0_p$, $\text{Var}\left(\varepsilon_i \right) = \sigma^2 > 0$ and with a positive definite matrix $\text{Cov}(x_i) = \Sigma_x$. Thus $z_i^T = \left(x_i^T, y_i \right)$ are i.i.d. copies of $z^T = \left(x^T, y \right)$ with expectation zero and covariance matrix

$$\text{Cov}_{\beta_0}(z) = \begin{pmatrix} \Sigma_x & \Sigma_x \beta_0 \\ \beta_0^T \Sigma_x & \beta_0^T \Sigma_x \beta_0 + \sigma^2 \end{pmatrix}.$$

The parameter of interest is $\beta_0 \in \mathbb{R}^p$. For an arbitrary parameter $\beta \in \mathbb{R}^p$ introduce $z\left(\beta \right) = y - \beta^T x$, then

$$\text{Cov}_{\beta_0}\left(x, z(\beta) \right) = \text{Cov}\left(x, \varepsilon + (\beta_0 - \beta)^T x \right) = \Sigma_x(\beta_0 - \beta).$$

Hence $\text{Cov}_{\beta_0}\left(x, z(\beta) \right)$ is zero if and only if $\beta = \beta_0$. Consider Kendall's tau as estimate of dependence between $z_i\left(\beta \right) = y_i - \beta^T x_i$ and $x_i = (x_{(k)i})$, $k = 1, \ldots, p$

$$\tilde{\tau}_k\left(\beta \right) = \frac{4}{n(n-1)} \sum_{i<j} \mathbf{I}\Big(\left(z_i\left(\beta \right) - z_j\left(\beta \right) \right) \left(x_{(k)i} - x_{(k)j} \right) < 0 \Big) - 1, \qquad (1)$$

where $\mathbf{I}\left((z_i(\beta) - z_j(\beta))(x_{(k)i} - x_{(k)j}) < 0\right)$ is 1 for $\left(z_i(\beta) - z_j(\beta)\right)\left(x_{(k)i} - x_{(k)j}\right) < 0$ and zero otherwise. In [12] the Theil-Sen estimate is introduced as the value of β where Kendall's tau is zero. Note, in [12] $p = 1$.

Using the Pearson's sample covariance we obtain the least squares estimate because

$$\widehat{\beta}_{lse} : \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \beta^T x_i\right) x_i = 0, \quad \widehat{\beta}_{lse} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i^T\right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i^T y_i. \quad (2)$$

2.2. Structural model

Consider now the linear model with errors in variables. The independent identically distributed random vectors $z_i^T = \left(x_i^T, y_i\right)^T$, $i = 1, \ldots, n$, $x_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}$ follow a linear errors-in-variables model

$$y_i = \beta_0^T \xi_i + \varepsilon_i, \tag{3}$$
$$x_i = \xi_i + \delta_i, \tag{4}$$

where ξ_i are i.i.d. random p-dimensional vectors with expectation zero and positive definite covariance matrix Σ_ξ. The errors ε_i in (3) are i.i.d. with expectation zero and variance σ^2, the errors δ_i in (4) are i.i.d. with expectation zero and covariance matrix $\sigma^2 I_p$, where I_p is the identity matrix of order p. Furthermore the errors ε_i, δ_i are uncorrelated. Summarizing the assumptions above we have that $z_i^T = \left(x_i^T, y_i\right)^T$ are i.i.d. copies of $z^T = (x^T, y)$ with $\epsilon^T = (\delta^T, \varepsilon)$ and

$$z = L_{\beta_0}\xi + \epsilon, \quad \text{and} \quad L_\beta = \begin{pmatrix} I_p \\ \beta^T \end{pmatrix}. \tag{5}$$

The $(p+1) \times 1$ random vector z has expectation zero and the covariance matrix

$$\Sigma(\beta_0) = L_{\beta_0}\Sigma_\xi L_{\beta_0}^T + \sigma^2 I_{p+1} \tag{6}$$

$$= \begin{pmatrix} \Sigma_\xi + \sigma^2 I_p & \Sigma_\xi \beta_0 \\ \beta_0^T \Sigma_\xi & \beta_0^T \Sigma_\xi \beta_0 + \sigma^2 \end{pmatrix}. \tag{7}$$

The parameter of interest is $\beta_0 \in \mathbb{R}^p$. Define

$$M_{zz} = \frac{1}{n} \sum_{i=1}^{n} z_i z_i^T = \begin{pmatrix} M_{xx} & M_{xy} \\ M_{yx} & M_{yy} \end{pmatrix},$$

with $M_{xx} = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T$, $M_{xy} = \frac{1}{n} \sum_{i=1}^{n} x_i y_i$, $M_{yx} = M_{xy}^T$ and $M_{yy} = \frac{1}{n} \sum_{i=1}^{n} y_i^2$ and $\lambda_1 = \lambda_{\min}(M_{zz})$. We will need the the following assumptions.

A1

$$M_{zz} = \frac{1}{n} \sum_{i=1}^{n} z_i z_i^T \quad \text{is positive definite.} \tag{8}$$

A2

$$\lambda_1 < \lambda_{\min}(M_{xx}). \tag{9}$$

These assumptions are not very restrictive. The matrix M_{zz} converges in probability to $\Sigma(\beta_0)$ with $\lambda_{\min}\Sigma(\beta_0) = \sigma^2$. The matrix M_{xx} converges in probability to $\Sigma_\xi + \sigma^2 I_p$ with $\lambda_{\min}(\Sigma_\xi + \sigma^2 I_p) = \lambda_{\min}(\Sigma_\xi) + \sigma^2$.

In difference to the regression model the $z(\beta_0) = y - \beta_0^T x$ and the observable variables x are not uncorrelated because of

$$\text{Cov}_{\beta_0}\big(x, z(\beta)\big) = \sigma^2 I_p + \Sigma_\xi (\beta_0 - \beta).$$

In model (5) we apply instead of x a different statistic $z_1(\beta)$ which is orthogonal to $z(\beta)$. For arbitrary $\beta \in \mathbb{R}^p$ we define the transformations

$$z_1(\beta) = x + \beta y \quad \text{and} \quad z_2(\beta) = y - \beta^T x. \tag{10}$$

For all $\beta \in \mathbb{R}^p$ it holds

$$L_\beta^\perp = \begin{pmatrix} -\beta \\ 1 \end{pmatrix} \quad \text{and} \quad L_\beta^T L_\beta^\perp = 0_p \tag{11}$$

and $z_1(\beta) = L_\beta^T z$ and $z_2(\beta) = \big(L_\beta^\perp\big)^T z$. Then we have the following statements.

THEOREM 1. *Under* (7) *and* (10) *it holds*

$$\text{Cov}_{\beta_0}\big(z_1(\beta), z_2(\beta)\big) = \big(I_p + \beta\beta_0^T\big) \Sigma_\xi (\beta_0 - \beta). \tag{12}$$

P r o o f. We have

$$\text{Cov}_{\beta_0}\big(z_1(\beta), z_2(\beta)\big) = \text{Cov}_{\beta_0}\big(L_\beta^T z, z^T L_\beta^\perp\big) = L_\beta^T \Sigma(\beta_0) L_\beta^\perp.$$

Using (6) and (11) we get

$$L_\beta^T \Sigma(\beta_0) L_\beta^\perp = L_\beta^T L_{\beta_0} \Sigma_\xi L_{\beta_0}^T L_\beta^\perp + \sigma^2 L_\beta^T L_\beta^\perp = L_\beta^T L_{\beta_0} \Sigma_\xi L_{\beta_0}^T L_\beta^\perp.$$

Further it holds $L_\beta^T L_{\beta_0} = I_p + \beta\beta_0^T$ and $L_{\beta_0}^T L_\beta^\perp = \beta_0 - \beta$. $\qquad\square$

THEOREM 2. *Under* (7) *and* (10) *it holds*

$$\text{Cov}_{\beta_0}\big(z_1(\beta), z_2(\beta)\big) = 0_p \Leftrightarrow \beta = \beta_0 \quad or \quad \beta^T \beta_0 = -1.$$

P r o o f. The condition $\text{Cov}_{\beta_0}\big(z_1(\beta), z_2(\beta)\big) = 0$ means

$$\big(I_p + \beta\beta_0^T\big) \Sigma_\xi (\beta_0 - \beta) = 0_p. \tag{13}$$

The p eigenvalues of $\big(I_p + \beta\beta_0^T\big)$ are $1, \ldots, 1, 1 + \beta^T\beta_0$. For $\beta^T \beta_0 \neq -1$ all eigenvalues are not zero, thus (13) has the unique solution $\beta = \beta_0$. Under $\beta^T \beta_0 = -1$ the matrix $I_p + \beta\beta_0^T$ has one eigenvalue equals zero and the corresponding eigenvector is proportional to β. Then $\beta_a = \big(a\Sigma_\xi^{-1} + I_p\big)^{-1}\beta_0$, where a is given such that $\beta_a^T \beta_0 = -1$, is the other unique solution of (13). $\qquad\square$

Geometrically, the condition $\beta^T \beta_0 = -1$ means that β and β_0 are oppositely oriented. We assume that we can exclude the case $\beta^T \beta_0 = -1$ by additional information. For the consistency of the rank estimate we require the stronger condition (15), see the following Lemma 3. We consider the correlations

$$\rho_{(k)\beta_0}(\beta) = \text{Cor}_{\beta_0}\big(z_{1(k)}(\beta), z_2(\beta)\big), \qquad k = 1, \ldots, p \qquad (14)$$

between $z_{1(k)}$ and z_2, where $z_{1(k)} = e_k^T z_1$ is the k'th component of z_1. e_k is the $p \times 1$ vector, whose components are zero except the k'th component which is equal to 1.

LEMMA 3. *Assume* (7), (10) *and*

$$\beta^T \beta_0 \geq -\frac{1}{2} + d, \qquad for \ some \quad d > 0 \qquad (15)$$

then it holds

$$\sum_{k=1}^{p} \rho_{(k)\beta_0}(\beta)^2 \geq \frac{c_d \left\| \beta_0 - \beta \right\|^2 + \left\| \beta \right\|^2 \left((\beta_0 - \beta)^T \Sigma_\xi \beta_0 \right)^2}{\lambda_{\max} \Sigma(\beta_0)^2 (1 + \left\| \beta \right\|^2)^2}$$

with $c_d = \min(1, 2d) \lambda_{\min} (\Sigma_\xi)^2$.

The proof of this lemma is technical and given in the Appendix.

3. Estimators

Using the statement of Theorem 1 we define estimators, which are based on a uniformly consistent estimate of the covariance between $z_1(\beta)$ and $z_2(\beta)$ or alternatively which are based on a monotone function of the correlations $\rho_{(k)\beta_0}(\beta)$, $k = 1, \ldots, p$ (Lemma 3).

3.1. Rank estimate

We introduce Kendall's tau $\widehat{\tau}_k(\beta)$ as an estimation of the dependence between the $z_{1(k)i} = e_k^T z_{1i}$ and z_{2i}, $i = 1, \ldots, n$,

$$\widehat{\tau}_k(\beta) = \frac{4}{n(n-1)} \sum_{i<j} \mathbf{I}\Big(\big(z_{1(k)i}(\beta) - z_{1(k)j}(\beta)\big)\big(z_{2i}(\beta) - z_{2j}(\beta)\big) < 0\Big) - 1.$$

$$(16)$$

More precisely, Kendall's tau $\widehat{\tau}_k(\beta)$ is a consistent estimate of

$$\tau_k(\beta) = P_{\beta_0}\Big(\big(z_{1(k)1}(\beta) - z_{1(k)2}(\beta)\big)\big(z_{21}(\beta) - z_{22}(\beta)\big) > 0\Big) - 1.$$

LEMMA 4. *Under* (7) *and* (10) *it holds*

$$\widehat{\tau}_k(\beta) \to^{P_{\beta_0}} \tau_k(\beta) \quad uniformly \ in \ \beta \quad and \quad k = 1, \ldots, p.$$

P r o o f. The statistic $\widehat{\tau}(\beta) + 1$ is a U statistic with kernel

$$h\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) = 2\mathbf{I}\big((x_1 - x_2)(y_1 - y_2) > 0\big).$$

Thus $\widehat{\tau}_k(\beta) + 1$ converges in probability P_{β_0} to

$$\theta_k(\beta) = 2P_{\beta_0}\Big(\big(z_{1(k)1}(\beta) - z_{1(k)2}(\beta)\big)\big(z_{21}(\beta) - z_{22}(\beta)\big) > 0\Big),$$

compare [16, Chapter 12, Example 12.5]. The proof of the uniform convergence follows the line of Step 2 in the proof in [1, p. 311]. □

Under normal distribution Greiner's relation $\tau_k(\beta) = \frac{2}{\pi} \arcsin\big(\rho_{(k)\beta_0}(\beta)\big)$ is valid, see for instance [4, p. 208]. These relation holds also for elliptical distributions.

DEFINITION 1. A p dimensional random vector X is called elliptical distributed with the parameters, μ, Σ, ϕ, where $\mu \in \mathbb{R}^p$ and Σ is a $p \times p$ nonnegative symmetric matrix, ϕ is a function from $[0, \infty)$ to $\mathbb{R}$ if and only if the characteristic function $\varphi_{X-\mu}$ of $X - \mu$ is of the form $\varphi_{X-\mu}(t) = \phi\big(t^T \Sigma t\big)$. We write $X \sim E_p(\mu, \Sigma, \phi)$.

If $X \sim E_p(\mu, \Sigma, \phi)$, then $AX \sim E_p\big(A\mu, A\Sigma A^T, \phi\big)$. Furthermore for absolute continuous r.v. $(X_1, X_2) \sim E_2(\mu, \Sigma, \phi)$ Kendall's population τ between (X_1, X_2) fulfills Greiner's relation $\tau = \frac{2}{\pi} \arcsin\big(\mathrm{Cor}(X_1, X_2)\big)$, compare [2, Theorem 3.1]. Further $X \sim E_p(\mu, \Sigma, \phi)$ and the $\mathrm{Cov}(X)$ exists, then $\mathrm{Cov}(X) = 2\phi'(0)\Sigma$. Introduce the distribution assumption.

E The random vector z is absolutely continuous and

$$z \sim E_{p+1}\big(0, a\Sigma(\beta_0), \phi\big), \qquad a > 0,$$

such that

$$\mathrm{Cov}(z) = \Sigma(\beta_0), \qquad \text{given in (7).} \tag{17}$$

DEFINITION 2. The Kendall's rank estimator $\widehat{\beta}_{rank}$ is a solution of

$$\widehat{\tau}_k\big(\widehat{\beta}_{rank}\big) = 0, \qquad k = 1, \dots, p \quad \text{and} \quad \widehat{\beta}_{rank}^T \beta_0 \neq -1.$$

Under additional assumptions we can prove the consistency of this rank estimator. The proof is technical and given in the Appendix.

THEOREM 5. *Assume E and that one of the following conditions is fulfilled.*

(i) *There exist constants $K < \infty$ and $d > 0$ such that*

$$\widehat{\beta}_{rank} \in \left\{\beta : \|\beta\|^2 \leq K, \ \beta^T \beta_0 \geq -\frac{1}{2} + d\right\}. \tag{18}$$

(ii) *There exist constants* $C = C(\Sigma_\xi, \beta_0) > 0$ *and* $d > 0$ *such that*

$$\widehat{\beta}_{rank} \in \left\{ \beta : \left(\beta^T \Sigma_\xi \beta_0 \right)^2 \geq C(\Sigma_\xi, \beta_0) \|\beta\|^2, \ \beta^T \beta_0 \geq -\frac{1}{2} + d \right\}. \quad (19)$$

Then it holds $\widehat{\beta}_{rank} \to^{P_{\beta_0}} \beta_0$.

Remark 1. The condition $\widehat{\beta}_{rank}^T \beta_0 \geq -\frac{1}{2} + d$ comes from Lemma 3. The main reason is that we have to exclude the second solution of (13), compare Theorem 2. The condition $\left(\beta^T \Sigma_\xi \beta_0 \right)^2 \geq C(\Sigma_\xi, \beta_0) \|\beta\|^2$ is fulfilled for $p = 1$ and $\beta_0 \neq 0$.

3.2. Total least squares estimate

In the structural model we can show a relation analogously to (2). The total least squares estimator is defined by

$$\widehat{\beta}_{tls} = \arg\min_\beta \frac{1}{n} \sum_{i=1}^{n} \min_\xi \|z_i - L_\beta \xi\|^2. \quad (20)$$

In [3, Chapter 2] it is derived that $\widehat{\beta}_{tls} = (M_{xx} - \lambda_1 I_p)^{-1} M_{xy}$.

THEOREM 6. *Under the conditions A1 and A2 the estimate* $\widehat{\beta}_{pear}$ *defined by*

$$\frac{1}{n} \sum_{i=1}^{n} z_{1i}(\widehat{\beta}_{pear}) z_{2i}(\widehat{\beta}_{pear}) = 0_p \quad and \quad \widehat{\beta}_{pear}^T M_{xy} > 0$$

is unique and $$\widehat{\beta}_{pear} = \left(M_{xx} - \lambda_1 I_p \right)^{-1} M_{xy}.$$

Thus $\widehat{\beta}_{pear}$ and the total least squares estimate coincides. The proof is given in the Appendix.

4. Simulation

In a very small simulation with $p = 1$ we compare the rank estimate $\widehat{\beta}_{rank}$ with $\widehat{\beta}_{tls}$ and the following estimates:

$$\widehat{\beta}_{naive} = \frac{M_{xy}}{M_{xx}}, \qquad \widehat{\beta}_{naive} \to^{P_{\beta_0}} \frac{\beta_0 \sigma_\xi^2}{\sigma_\xi^2 + \sigma^2},$$

$$\widehat{\beta}_{inv} = \frac{M_{yy}}{M_{xy}}, \qquad \widehat{\beta}_{inv} \to^{P_{\beta_0}} \frac{\beta_0 \sigma_\xi^2 + \sigma^2}{\beta_0 \sigma_\xi^2}.$$

The simulation is done in a bad case situation, small sample size (n=10) and high error variances $\sigma^2 = 3$ in relation to the variance of unobservables $\sigma_\xi^2 = 3$. The samples are calculated from (3) and (4), where $(\xi_i, \varepsilon_i, \delta_i)$ are independently generated t-distributed random variables with three degrees of freedom. The tls

estimator is consistent and has optimality properties, see [11]. But there exists no moments of the tls estimates. The tls estimator is often unstable, the same is valid for the iteration procedure needed for the calculation of the rank estimator. But there are still reserves in stabilizing the numerical procedure of the rank estimator (stopping earlier, searching for other start values and so on). One conclusion is clear the tls and the rank estimates are consistent, the naive estimate and the inverse estimate are not. The naive estimate is more concentrated but around the wrong value.

Simulation with n=10, true beta=2, rep=50
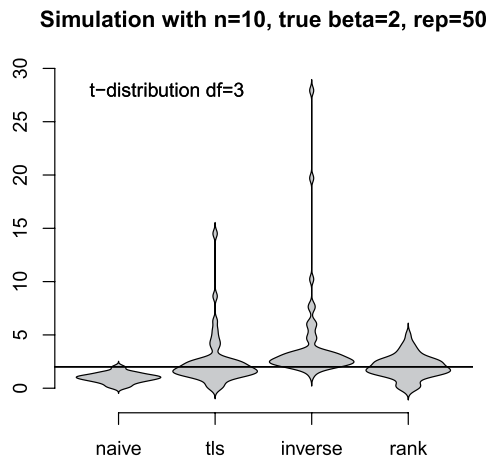
5. Appendix

Proof of Lemma 3. Consider

$$\rho_{(k)\beta_0}(\beta)^2 = \frac{\mathrm{Cov}(z_{1(k)}(\beta), z_2(\beta))^2}{\mathrm{Var}(z_{1(k)}(\beta))\,\mathrm{Var}(z_2(\beta))}.$$

Remind

$$z_{1(k)}(\beta) = e_k^T L_\beta^T z \quad \text{and} \quad z_2(\beta) = \left(L_\beta^\perp\right)^T z, \quad \mathrm{Cov}_{\beta_0}(z) = \Sigma(\beta_0) \quad (21)$$

and

$$\lambda_{\max}\left(L_\beta^T L_\beta\right) = 1 + \|\beta\|^2 \quad \text{and} \quad \left\|L_\beta^\perp\right\|^2 = 1 + \|\beta\|^2.$$

Then

$$\mathrm{Var}\left(z_{1(k)}(\beta)\right) \leq \lambda_{\max}\Sigma(\beta_0)\left\|L_\beta e_k\right\|^2 \leq \lambda_{\max}\Sigma(\beta_0)\left(1 + \|\beta\|^2\right)$$

and

$$\mathrm{Var}\left(z_2(\beta)\right) \leq \lambda_{\max}\Sigma(\beta_0)\left\|L_\beta^\perp\right\|^2 \leq \lambda_{\max}\Sigma(\beta_0)\left(1 + \|\beta\|^2\right).$$

Hence

$$\sum_{k=1}^{p} \rho_{(k)\beta_0}(\beta)^2 \geq \left(\lambda_{\max}\Sigma(\beta_0)\left(1 + \|\beta\|^2\right)\right)^{-2} S(\beta).$$

with

$$S(\beta) = \sum_{k=1}^{p} \text{Cov}\left(z_{1(k)}(\beta), z_2(\beta)\right)^2.$$

From Theorem 1 we get

$$S(\beta) = \sum_{k=1}^{p} \left(e_k^T \left(I_p + \beta\beta_0^T\right) \Sigma_\xi \left(\beta_0 - \beta\right)\right)^2.$$

Because $\sum_{k=1}^{p} e_k e_k^T = I_p$, it holds

$$S(\beta) = (\beta_0 - \beta)^T \Sigma_\xi \left(I_p + \beta_0\beta^T\right) \left(I_p + \beta\beta_0^T\right) \Sigma_\xi (\beta_0 - \beta) = S_1(\beta) + S_2(\beta)$$

with

$$S_1(\beta) = (\beta_0 - \beta)^T \Sigma_\xi \Sigma_\xi (\beta_0 - \beta) + 2(\beta_0 - \beta)^T \Sigma_\xi \beta\beta_0^T \Sigma_\xi (\beta_0 - \beta)$$

and

$$S_2(\beta) = \|\beta\|^2 \left((\beta_0 - \beta)^T \Sigma_\xi \beta_0\right)^2 \geq 0.$$

The eigenvalues of $I + 2\beta_0\beta^T$ are $1, \ldots, 1, 1 + 2\beta_0^T\beta$. Thus

$$\begin{aligned} S_1(\beta) &\geq \lambda_{\min}\left(I + 2\beta_0\beta^T\right) \|\Sigma_\xi (\beta_0 - \beta)\|^2 \\ &\geq \min(1, 2d)\lambda_{\min}\left(\Sigma_\xi\right)^2 \|\beta_0 - \beta\|^2. \end{aligned}$$

$\square$

Proof of Theorem 5. Because of (21) and $\text{E}\, z_{1(k)}$ and z_2 are elliptically distributed and Greiner's relation holds. Hence

$$\tau_k(\beta)^2 \geq \frac{1}{\pi^2} \left|\rho_{(k)\beta_0}(\beta)\right|^2 \qquad \text{for all} \quad k.$$

Applying Lemma 3 on $C(\beta) = \sum_{k=1}^{p} \tau_k(\beta)^2$ and remind $\widehat{\tau}_k\left(\widehat{\beta}_{rank}\right) = 0$ for all k, we obtain

$$\begin{aligned} C\left(\widehat{\beta}_{rank}\right) &= \sum_{k=1}^{p} \left(\tau_k\left(\widehat{\beta}_{rank}\right) - \widehat{\tau}_k\left(\widehat{\beta}_{rank}\right)\right)^2 \\ &\geq c\left(1 + \left\|\widehat{\beta}_{rank}\right\|^2\right)^{-2} \left\|\beta_0 - \widehat{\beta}_{rank}\right\|^2. \end{aligned}$$

Consider an arbitrary $\varepsilon > 0$ and an arbitrary constant K. Then

$$\left\|\widehat{\beta}_{rank} - \beta_0\right\|^2 > \varepsilon \quad \text{and} \quad \left\|\widehat{\beta}_{rank}\right\| \leq K$$

imply that there exists a constant c_K such that $C(\widehat{\beta}_{rank}) > \varepsilon c_K$. Consider now $\left\|\widehat{\beta}_{rank}\right\| > K$ and condition (19). Apply Lemma 3 and take as leading term now $S_2(\beta) = \|\beta\|^2 \left((\beta_0 - \beta)^T \Sigma_\xi \beta_0\right)^2$, then

$$
\begin{aligned}
S_2(\beta) &\geq \|\beta\|^4 \left(\frac{(\beta^T \Sigma_\xi \beta_0)^2}{\|\beta\|^2} + \frac{(\beta_0^T \Sigma_\xi \beta_0)^2}{\|\beta\|^2} - 2 \frac{\beta_0^T \Sigma_\xi \beta_0 \beta^T \Sigma_\xi \beta_0}{\|\beta\|^2} \right) \\
&\geq \|\beta\|^4 \left(C(\Sigma_\xi, \beta_0) - 2\beta_0^T \Sigma_\xi \beta_0 \frac{|\beta^T \Sigma_\xi \beta_0|}{\|\beta\|^2} \right).
\end{aligned}
$$

Because of $\left|\beta^T \Sigma_\xi \beta_0\right| \leq \|\beta\| \|\beta_0\| \lambda_{\max}(\Sigma_\xi)$ there exists a constant $c(\Sigma_\xi, \beta_0)$ such that

$$
S_2(\beta) \geq \|\beta\|^4 C(\Sigma_\xi, \beta_0) - \|\beta\|^3 c(\Sigma_\xi, \beta_0).
$$

Applying Lemma 3 we get

$$
\min_{\|\beta\| > K} C(\beta) \geq \text{const } F(K),
$$

where $F(K) = K^4(1 - cK^{-1})/(1 + K^2)^2$. $F(K)$ is monotone and converges for $K \to \infty$ to 1. Hence for sufficiently large K there exists a constant $C_0 > 0$ such that $\min_{\|\beta\| > K} C(\beta) \geq C_0 > 0$. Using the inequality

$$
P\left(\sum_{k=1}^p |T_k| > \epsilon \right) \leq \sum_{k=1}^p P\left(|T_k| > \frac{\epsilon}{p} \right)
$$

we obtain

$$
\begin{aligned}
P_{\beta_0}\left(\left\|\beta_0 - \widehat{\beta}_{rank}\right\|^2 > \varepsilon \right) &\leq \sum_{k=1}^p P_{\beta_0}\left(\sup_{\|\beta\| \leq K, \|\beta_0 - \beta\| > \varepsilon} |\tau_k(\beta) - \widehat{\tau}_k(\beta)|^2 > \varepsilon \frac{c_K}{p} \right) \\
&\quad + \sum_{k=1}^p P_{\beta_0}\left(\sup_{\|\beta\| > K} |\tau_k(\beta) - \widehat{\tau}_k(\beta)|^2 > \frac{C_0}{p} \right).
\end{aligned}
$$

Then from Lemma 5 it follows that $P_{\beta_0}\left(\left\|\beta_0 - \widehat{\beta}_{rank}\right\|^2 > \varepsilon \right)$ converges to zero. $\square$

P r o o f   o f   T h e o r e m 6. From $L_\beta^T M_{zz} L_\beta^\perp = 0_p$ it follows that

$$
L_\beta L_\beta^T M_{zz} L_\beta^\perp = 0_p. \tag{22}
$$

The $(p+1) \times (p+1)$ symmetric matrix $L_\beta L_\beta^T$ has the eigenvalues $0, 1, \ldots, 1$, $1 + \|\beta\|^2$ and the eigenvector belonging to the eigenvalue 0 is proportional to $L_\beta^\perp$. Hence the vector $M_{zz} L_\beta^\perp$ belongs to the space spanned by $L_\beta^\perp$, thus there exists a scalar a such that $M_{zz} L_\beta^\perp = a L_\beta^\perp$. That means $L_\beta^\perp$ is proportional to an eigenvector belonging to one of the eigenvalues λ_i of M_{zz}. Thus

$$
M_{zz} L_\beta^\perp = \lambda_i L_\beta^\perp \quad \text{and} \quad (M_{zz} - \lambda_i I_{p+1}) L_\beta^\perp = 0_{p+1}.
$$

Because of (11) and the block structure of M_{zz} it holds

$$-M_{xx}\beta + M_{xy} + \lambda_i\beta = 0_p \quad \text{and} \quad -M_{yx}\beta + M_{yy} - \lambda_i = 0.$$

Hence the solutions of (22) are of the form

$$\beta = (M_{xx} - \lambda_i I_p)^- M_{xy}, \qquad i = 1, \dots, p+1,$$

where λ_i is an eigenvalue of M_{zz}. Consider the second condition $\beta^T M_{xy} > 0$ in order to specify the solutions. The second condition is equivalent to

$$M_{xy}^T (M_{xx} - \lambda_i I_p)^- M_{xy} > 0. \tag{23}$$

We have

$$M_{xy}^T (M_{xx} - \lambda_i I_p)^- M_{xy} \le \lambda_{\max}\big((M_{xx} - \lambda_i I_p)^-\big) \|M_{xy}\|^2.$$

The definition of a generalized inverse and the relation

$$\lambda(M_{xx} - \lambda_i I_p) = \lambda(M_{xx}) - \lambda_i$$

imply that the eigenvalues of $(M_{xx} - \lambda_i I_p)^-$ are

$$\lambda(M_{xx} - \lambda_i I_p)^- = \begin{cases} (\lambda(M_{xx}) - \lambda_i)^{-1} & \text{if} \quad \lambda(M_{xx}) \ne \lambda_i, \\ 0 & \text{if} \quad \lambda(M_{xx}) = \lambda_i \end{cases}$$

(compare for instance [10]). Thus for i with $\lambda(M_{xx}) = \lambda_i$ it holds

$$M_{xy}^T (M_{xx} - \lambda_i I_p)^- M_{xy} \le 0.$$

We can exclude these solutions, because of (23). For $i > 1$ we apply the interlacing inequality between the eigenvalues $\lambda_1 \le \cdots \le \lambda_{p+1}$ of M_{zz} and the eigenvalues $\alpha_1 \le \cdots \le \alpha_p$ of M_{xx} (compare for instance [15]) $\lambda_j \le \alpha_j \le \lambda_{j+1}$, $j = 1, \dots, p$. Hence for $i > 1$ with $\lambda(M_{xx}) \ne \lambda_i$ it holds

$$\lambda_{\min}(M_{xx}) - \lambda_i = \alpha_1 - \lambda_i \le \alpha_1 - \lambda_2 \le 0$$

and

$$M_{xy}^T (M_{xx} - \lambda_i I_p)^- M_{xy} \le \frac{1}{\lambda_{\min}(M_{xx}) - \lambda_i} \|M_{xy}\|^2 \le 0.$$

That is a contradiction to (23) and we can exclude $i > 1$. Under the assumption A2 for $i = 1$ it holds

$$\lambda_{\min}(M_{xx} - \lambda_1 I_p) = \lambda_{\min}(M_{xx}) - \lambda_1 = \alpha_1 - \lambda_1 > 0$$

and $M_{xx} - \lambda_1 I_p$ is positive definite and condition (23) is fulfilled. $\qquad \square$

REFERENCES

[1] HAN, A. K.: *Non-parametric analysis of a generalized regression model*, J. Econom. **35** (1987), 303–316.

[2] FANG, H.-B.—FANG, K.-T.: *The meta-elliptical distributions with given marginals*, J. Multivariate Anal. **82** (2002), 1–16.

[3] FULLER, W. A.: *Measurement Error Models.* John Wiley & Sons, New York, 1987.

[4] GIBBONS, J. D.: *Nonparametric Statistical Inference.* McGraw-Hill Book Comp., New York, 1971.

[5] JAECKEL, L. A.: *Estimating regression coefficients by minimizing the dispersion of the residuals*, Ann. Math. Stat. **43** (1972), 1449–1458.

[6] JUREČKOVÁ, J.: *Asymptotic linearity of a rank statistic in regression parameter*, Ann. Math. Stat. **40** (1969), 1889–1900.

[7] JUREČKOVÁ, J.: *Nonparametric estimate of regression coefficients*, Ann. Math. Stat. **42** (1971), 1328–1338.

[8] JUREČKOVÁ, J.—PICEK, J.—SALEH, A. K. MD. E.: *Rank tests and regression rank scores tests in measurement error models*, Comput. Stat. Data Anal., 2012 (in press).

[9] JUREČKOVÁ, J.—SEN, PRANAB, K.: *Robust Statistical Procedures*, John Wiley & Sons, Inc., New York, 1996.

[10] MARDIA, K. V.—KENT, J. T.—BIBBY, J. M.: *Multivariate Analysis*, Academic Press, London, 1979.

[11] NUSSBAUM, M.: *An asymptotic minimax risk bound for estimation of a linear functional relationship*, J. Multivariate Anal. **14** (1984), 300–314.

[12] SEN, P. K.: *Estimates of the regression coefficient based on Kendall's tau*, J. Amer. Statist. Assoc. **63** (1968), 1379–1389.

[13] SEN, P. K.—SALEH, A. K. MD. E.: *The Theil–Sen estimator in a measurement error perspective.* In: Nonparametrics and robustness in modern statistical inference and time series analysis: Festschrift for Jana Jurečková, Inst. Math. Stat. Collect., Vol. 7, Inst. Math. Statist., Beachwood, OH, 2010, pp. 224–235.

[14] PE, TH.: *Analysis of functional relationships: old and new methods*, Bull. Austral. Math. Soc. **14** (1976), 145–147.

[15] THOMPSON, R. C.: *Principal submatrices. VIII. Principal sections of a pair of forms*, Rocky Mt. J. Math. **2** (1972), 97–110.

[16] VAN DER VAART, A. W.: *Asymptotic Statistics*, in: Stat. Probab. Math., Vol. 3, Cambridge University Press, Cambridge, 1998.

Department of Mathematics
Uppsala University
P.O. Box 480
SE–75106 Uppsala
SWEDEN
E-mail: zwanzig@math.uu.se