Теорія Ймовір. та Матем. Статист.
Вип. 97, 2017

Theor. Probability and Math. Statist.
No. 97, 2018, Pages 45–55
https://doi.org/10.1090/tpms/1047
Article electronically published on February 21, 2019

CONSISTENCY OF THE ORTHOGONAL REGRESSION ESTIMATOR IN AN IMPLICIT LINEAR MODEL WITH ERRORS IN VARIABLES

UDC 519.21

O. O. DASHKOV AND A. G. KUKUSH

Abstract. An implicit linear regression model with errors in variables is studied for which the true points belong to a certain hyperplane in Euclidean space and the joint covariance matrix of errors is proportional to the unit matrix. The orthogonal regression estimator for this hyperplane is considered. Some sufficient conditions for the consistency as well as for the strong consistency are given. Some applications to the explicit multiple regression model with a free term and errors in variables are shown.

1. Introduction

The multiple functional linear regression model with a free term and errors in variables is described by the following equalities:

$$\begin{cases} y_i = b_0 + x^\top \xi_i + \varepsilon_i, \\ x_i = \xi_i + \delta_i. \end{cases} \tag{1}$$

Here $\xi_i \in \mathbb{R}^{m-1}$ are unknown non-random vectors, ε_i and δ_i are random errors, and $b_0 \in \mathbb{R}$ and $x \in \mathbb{R}^{m-1}$ are regression parameters that need to be estimated from observations (y_i, x_i), $i = 1, \dots, n$. Similar models of observations occur often in econometrics and in signal processing. Here and in what follows all vectors are column vectors.

This model is reduced usually to a linear regression without free term and then the least square estimator is considered for the reduced model (total least squares estimator). In particular, this estimator is studied in the papers [1–3, 5], where some sufficient conditions for the weak as well as for strong consistency are obtained for the total least squares estimator under various assumptions concerning the model of observations. The least restrictive conditions for the case considered in the current paper are obtained in [7].

In order to cancel the free term in the model (1) one introduces an additional regressor which is constant and equal to unity and is observed without errors. Such an approach has a drawback, since the regressors in the resulting model are unequal. To avoid this problem, we consider an implicit linear model for observations in the Euclidean space $\mathbb{R}^m$, $m \geq 2$, where the hidden variables belong to a certain hyperplane and this hyperplane needs to be estimated.

The most frequently used estimator for such an implicit model is the orthogonal regression estimator. The orthogonal regression estimator defines a hyperplane with the least sum of distances from observations. Since all variables are equally treated in this

---

2010 *Mathematics Subject Classification.* Primary 62J05; Secondary 62H12, 65F20.

Key words and phrases. Consistency of an estimator, multiple regression model with errors in variables, implicit linear regression model, orthogonal regression model, total least squares estimator.

case, one can apply some geometrical intuition and reasoning when studying properties of the estimator.

Using all benefits of an implicit model, we are able to find sufficient conditions for the consistency as well as for the strong consistency of the orthogonal regression estimator. These conditions are imposed on eigenvalues of the sample covariance matrix of hidden variables. Note that the conditions do not follow from the corresponding conditions for consistency in [7] in the case of the reduced model without free term.

The paper is organized as follows. We describe the implicit linear model in Section 2 for observations in the Euclidean space. In Section 2, the orthogonal regression estimator is introduced and the assumptions concerning the errors of observations are stated. Sections 3 and 4 contain sufficient conditions for the consistency and strong consistency, respectively, of the orthogonal regression estimator. These results are used in Section 5 for the total least squares estimator for the multiple linear regression with a free term and errors in variables. In Section 6, the multiple model with a free term is reduced to a model without free term. We also compare our results with known theorems concerning the total least squares estimators in Section 6. In Section 7, we briefly discuss a possible direction for further investigations.

Throughout this paper the bar above a letter, say $\bar{u}$, means the averaging of the first n members of a sequence u of numbers, vectors, or matrices.

A number in brackets written as an upper index, say $\bar{\eta}^{(n)}$, denotes the index of a member of an averaged sequence $\bar{\eta}$. If this index is absent for a member of an averaged sequence, then the index of this member equals n.

2. The estimator and assumptions concerning the model

Consider the following implicit regression linear model of observations in the Euclidean space $\mathbb{R}^m$, $m \geq 2$:

$$\begin{cases} z_i = \eta_i + \gamma_i, \\ (\eta_i, \tau) = d. \end{cases} \tag{2}$$

Here z_i denote the observed vectors in $\mathbb{R}^m$, η_i are hidden variables that belong to the hyperplane

$$\Gamma_{\tau d} = \{u \in \mathbb{R}^m \colon (u, \tau) = d\}, \tag{3}$$

τ is the unit vector being normal to the hyperplane, $d \in \mathbb{R}$, and γ_i are random errors of observations. Using the observations z_i, $i = 1, \dots, n$, we need to estimate the hyperplane (3).

Below we list the assumptions concerning model (2).

(i) The vectors γ_i, $i \geq 1$, are jointly independent and have zero mean.
(ii) The vectors γ_i, $i \geq 1$, have identical covariance matrix $\mathbf{cov}(\gamma_i) = \sigma^2 I_m$, where $\sigma^2 > 0$ is unknown.

For given sets u_i, $i = 1, \dots, n$, and v_i, $i = 1, \dots, n$, of vectors of the space $\mathbb{R}^m$, we consider the sample covariance matrix

$$S_{uv} = \frac{1}{n} \sum_{i=1}^{n} (u_i - \bar{u})(v_i - \bar{v})^\top, \qquad \bar{u} = \frac{1}{n} \sum_{i=1}^{n} u_i, \qquad \bar{v} = \frac{1}{n} \sum_{i=1}^{n} v_i.$$

The orthogonal regression estimator of parameters τ and d in model (2) is defined by the pair $\hat{\tau}$ and $\hat{d}$ for which the goal function

$$Q(\tau, d) = \sum_{i=1}^{n} \rho^2(z_i, \Gamma_{\tau d}), \qquad \|\tau\| = 1,\ d \in \mathbb{R},$$

attains its minimum, where $\rho(z,\Gamma)$ denotes the distance between the point z and hyperplane Γ.

It is shown in [6] that the orthogonal regression estimator $\hat{\tau}$ is a normalized eigenvector of the function S_{zz} corresponding to the minimal eigenvalue $\lambda_{\min}(S_{zz})$ and that

$$\hat{d} = (\bar{z}, \hat{\tau}).$$

Given a symmetric matrix $S_{\eta\eta}$, we write its eigenvalues with their multiplicities in the ascending order:

$$\lambda_{\min}(S_{\eta\eta}) \le \lambda_2(S_{\eta\eta}) \le \cdots \le \lambda_{\max}(S_{\eta\eta}).$$

In what follows, Σ means $\sigma^2 I_m = \mathbf{cov}(\gamma_i)$.

3. Sufficient conditions for consistency of the orthogonal regression estimator

Theorem 1. *Assume that conditions* (i) *and* (ii) *hold for models* (2)–(3). *We further assume that, for some* $1 \le r \le 2$,

$$\sup_{i\ge 1} \mathsf{E}\,\|\gamma_i\|^{2r} < \infty,$$

$$\frac{\lambda_{\max}\left(S_{\eta\eta}^{(n)}\right)}{n\lambda_2^2\left(S_{\eta\eta}^{(n)}\right)} \to 0, \qquad n \to \infty,$$

$$n^{1-1/r}\lambda_2\left(S_{\eta\eta}^{(n)}\right) \to \infty, \qquad n \to \infty.$$

Then the estimator $\hat{\tau}$ *is consistent, that is,*

$$\min\bigl\{\|\hat{\tau}-\tau\|, \|\hat{\tau}+\tau\|\bigr\} \xrightarrow{\mathsf{P}} 0$$

as $n \to \infty$.

Proof. Since the orthogonal regression estimator is equivariant with respect to orthogonal transformations (this property of the estimator is discussed in [6]), we assume that $\tau = (1,0,\ldots,0)^\top$ without loss of generality. In this case,

$$S_{\eta\eta}^{(n)} = \begin{pmatrix} 0 & 0_{1\times(m-1)} \\ 0_{(m-1)\times 1} & A_n \end{pmatrix},$$

where A_n is a square matrix of order $m-1$, since the vectors $\eta_i - \bar{\eta}^{(n)}$ belong to a hyperplane being perpendicular to the vector τ. Moreover,

$$\lambda_{\min}(A_n) = \lambda_2\left(S_{\eta\eta}^{(n)}\right).$$

Then

$$S_{zz}\hat{\tau} = \lambda_{\min}(S_{zz})\hat{\tau}, \qquad S_{zz} = S_{\eta\eta} + S_{\eta\gamma} + S_{\gamma\eta} + S_{\gamma\gamma}. \tag{4}$$

Since S_{zz} is a symmetric matrix,

$$\lambda_{\min}(S_{zz}) \le (S_{zz}\tau,\tau) = (S_{\gamma\gamma}\tau,\tau) = \sigma^2 + \bigl((S_{\gamma\gamma}-\Sigma)\tau,\tau\bigr). \tag{5}$$

On the other hand,

$$\lambda_{\min}(S_{zz}) = (S_{zz}\hat{\tau},\hat{\tau}) = (S_{\eta\eta}\hat{\tau},\hat{\tau}) + (S_{\eta\gamma}\hat{\tau},\hat{\tau}) + (S_{\gamma\eta}\hat{\tau},\hat{\tau}) + (S_{\gamma\gamma}\hat{\tau},\hat{\tau}).$$

Further, the matrix $S_{\eta\eta}$ is non-negative definite and $\|\hat{\tau}\| = 1$, whence

$$\lambda_{\min}(S_{zz}) - \sigma^2 \ge -(\|S_{\eta\gamma}\| + \|S_{\gamma\eta}\| + \|S_{\gamma\gamma}-\Sigma\|). \tag{6}$$

It follows from (5) and (6) that

$$\bigl|\lambda_{\min}(S_{zz}) - \sigma^2\bigr| \le \|S_{\eta\gamma}\| + \|S_{\gamma\eta}\| + \|S_{\gamma\gamma}-\Sigma\|. \tag{7}$$

We rewrite the first equality in (4) as follows:

$$S_{\eta\eta}\hat{\tau} = \Big(\left(\lambda_{\min}(S_{zz}) - \sigma^2\right) I_m - S_{\eta\gamma} - S_{\gamma\eta} - (S_{\gamma\gamma} - \Sigma)\Big)\hat{\tau}. \tag{8}$$

Let

$$B_n = \begin{pmatrix} 0 & 0_{1\times(m-1)} \\ 0_{(m-1)\times 1} & A_n^{-1} \end{pmatrix}$$

and let P_2 be the matrix of the orthogonal projector to the hyperplane $\{\tau\}^{\perp}$. Then

$$S_{\eta\eta}^{(n)} B_n = B_n S_{\eta\eta}^{(n)} = \begin{pmatrix} 0 & 0 \\ 0 & I_{m-1} \end{pmatrix} = P_2.$$

Multiplying from the left on both sides of equality (8) by B_n, we obtain

$$P_2\hat{\tau} = B_n\Big(\left(\lambda_{\min}(S_{zz}) - \sigma^2\right) I_m - S_{\eta\gamma} - S_{\gamma\eta} - \left(S_{\gamma\gamma} - \sigma^2 I_m\right)\Big)\hat{\tau}.$$

Taking into account inequality (7) we conclude that

$$\|P_2\hat{\tau}\| \le 2\|B_n\| \cdot \left(\|S_{\eta\gamma}\|_F + \|S_{\gamma\eta}\|_F + \|S_{\gamma\gamma} - \sigma^2 I_m\|_F\right), \tag{9}$$

since the operator norm of a matrix does not exceed its Frobenius norm. Here and in what follows, $\|A\|_F$ denotes the Frobenius norm of a matrix defined as the square root of the sum of squares of all matrix entries. Recall that the operator norm of a matrix is defined as the norm of the corresponding operator in an Euclidean space.

Since $\|\hat{\tau}\| = 1$, we have $\|P_2\hat{\tau}\|^2 + (\hat{\tau}, \tau)^2 = 1$ and hence

$$\min\big(\|\hat{\tau} - \tau\|, \|\hat{\tau} + \tau\|\big) = \sqrt{\left(1 - \sqrt{1 - \|P_2\hat{\tau}\|^2}\right)^2 + \|P_2\hat{\tau}\|^2}. \tag{10}$$

Therefore we only need to prove the convergence $P_2\hat{\tau} \xrightarrow{\mathsf{P}} 0$, $n \to \infty$. In turn, this follows from the convergence $\|B_n\| \cdot \big\|S_{\gamma\eta}^{(n)}\big\|_F \xrightarrow{\mathsf{P}} 0$ and

$$\|B_n\| \cdot \left\|S_{\gamma\gamma}^{(n)} - \Sigma\right\|_F \xrightarrow{\mathsf{P}} 0, \qquad n \to \infty,$$

in view of inequality (9).

Since B_n is a block diagonal matrix and since A_n^{-1} is a symmetric matrix,

$$\|B_n\| = \max\left(0, \left\|A_n^{-1}\right\|\right) = \lambda_{\max}\left(A_n^{-1}\right) = \frac{1}{\lambda_2\left(S_{\eta\eta}^{(n)}\right)}. \tag{11}$$

To estimate $\|S_{\gamma\eta}\|_F$, we apply the Rosenthal inequality for $2r \ge 2$ (written in a special form suggested in [7]):

$$\begin{aligned} \mathsf{E}\left\|S_{\gamma\eta}^{(n)}\right\|_F^{2r} &= \frac{1}{n^{2r}}\,\mathsf{E}\left\|\sum_{i=1}^n \gamma_i(\eta_i - \bar{\eta})^\top\right\|_F^{2r} \\ &\le \frac{\alpha}{n^{2r}}\sum_{i=1}^n \mathsf{E}\,\|\gamma_i\|^{2r}\,\|\eta_i - \bar{\eta}\|^{2r} + \frac{\beta}{n^{2r}}\left(\sum_{i=1}^n \mathsf{E}\,\|\gamma_i\|^2\,\|\eta_i - \bar{\eta}\|^2\right)^r. \end{aligned} \tag{12}$$

Further,

$$\mathsf{E}\,\|\gamma_i\|^{2r} \le \sup_{j\ge 1}\mathsf{E}\,\|\gamma_j\|^{2r}, \qquad \left(\mathsf{E}\,\|\gamma_i\|^2\right)^r \le \sup_{j\ge 1}\mathsf{E}\,\|\gamma_j\|^{2r}, \qquad i\ge 1,$$

$$\sum_{i=1}^n \|\eta_i-\bar\eta\|^2 = \operatorname{tr}\left(\sum_{i=1}^n(\eta_i-\bar\eta)(\eta_i-\bar\eta)^\top\right) = n\cdot\operatorname{tr}\left(S_{\eta\eta}^{(n)}\right) \le nm\lambda_{\max}\left(S_{\eta\eta}^{(n)}\right),$$

$$\sum_{i=1}^n \|\eta_i-\bar\eta\|^{2r} \le \left(\sum_{i=1}^n\|\eta_i-\bar\eta\|^2\right)^r \le n^r m^r\lambda_{\max}\left(S_{\eta\eta}^{(n)}\right)^r.$$

Thus

$$\mathsf{E}\,\|S_{\gamma\eta}\|_F^{2r} = \frac{O\left(\lambda_{\max}\left(S_{\eta\eta}^{(n)}\right)^r\right)}{n^r}, \qquad n\to\infty. \tag{13}$$

Then (11), (13), and assumptions of the theorem imply that

$$\|B_n\|\cdot\left\|S_{\gamma\eta}^{(n)}\right\|_F \xrightarrow{\mathsf{P}} 0, \qquad n\to\infty. \tag{14}$$

Now we estimate

$$\mathsf{E}\,\|S_{\gamma\gamma}-\Sigma\|_F^r \le \left(\left(\mathsf{E}\left\|\overline{\gamma\gamma^\top}-\Sigma\right\|_F^r\right)^{1/r} + \left(\mathsf{E}\left\|\bar\gamma\,\overline{\gamma^\top}\right\|_F^r\right)^{1/r}\right)^r. \tag{15}$$

By assumptions of Theorem 1, the sequence $\left\{\mathsf{E}\left\|\gamma_i\gamma_i^\top-\Sigma\right\|^r,\, i\ge 1\right\}$ is bounded and the random matrices $\left\{\gamma_i\gamma_i^\top-\Sigma, i\ge 1\right\}$ are jointly independent. Hence one can use the Rosenthal inequality for $1\le r\le 2$ (again in a special form as in [7]):

$$\begin{aligned}\mathsf{E}\left\|\overline{\gamma\gamma^\top}-\Sigma\right\|_F^r &= \frac{1}{n^r}\,\mathsf{E}\left\|\sum_{i=1}^n\left(\gamma_i\gamma_i^\top-\Sigma\right)\right\|_F^r\\ &\le \frac{\alpha}{n^r}\sum_{i=1}^n\mathsf{E}\left\|\gamma_i\gamma_i^\top-\Sigma\right\|_F^r = O\left(n^{1-r}\right).\end{aligned} \tag{16}$$

In addition,

$$\begin{aligned}\mathsf{E}\left\|\bar\gamma\,\overline{\gamma^\top}\right\|_F^r &= \mathsf{E}\left\|\bar\gamma^{(n)}\right\|^{2r} = \frac{1}{n^{2r}}\,\mathsf{E}\left\|\sum_{i=1}^n\gamma_i\right\|^{2r}\\ &\le \frac{\alpha}{n^{2r}}\sum_{i=1}^n\mathsf{E}\,\|\gamma_i\|^{2r} + \frac{\beta}{n^{2r}}\left(\sum\mathsf{E}\,\|\gamma_i\|^2\right)^r = O\left(n^{-r}\right).\end{aligned} \tag{17}$$

Therefore inequality (15) implies that $\|B_n\|\cdot\left\|S_{\gamma\gamma}^{(n)}-\Sigma\right\| \xrightarrow{\mathsf{P}} 0$, $n\to\infty$, and this completes the proof of the theorem. □

Theorem 2. *Let all assumptions of Theorem 1 hold and let the sequence $\bar\eta^{(n)}$, $n\ge 1$, be bounded. Then the orthogonal regression estimator $\left(\hat\tau;\hat d\right)$ is consistent, that is,*

$$\min\left\{\|\hat\tau-\tau\|+\left|\hat d-d\right|,\ \|\hat\tau+\tau\|+\left|\hat d+d\right|\right\} \xrightarrow{\mathsf{P}} 0 \tag{18}$$

as $n\to\infty$.

Remark. The convergence in (18) is explained by the property that the pair $\tau=\tau_0$, $d=d_0$ determines the same hyperplane as the pair $\tau=-\tau_0$, $d=-d_0$.

Proof. By the properties of the orthogonal regression estimator, $\hat{d} = (\hat{\tau}; \bar{z})$, whence $d = (\tau; \bar{\eta})$. Thus

$$\|\hat{\tau} - \tau\| + |\hat{d} - d| = \|\hat{\tau} - \tau\| + |(\hat{\tau}, \bar{z}) - (\tau, \bar{\eta})| = \|\hat{\tau} - \tau\| + |(\hat{\tau}, \bar{\gamma}) - (\hat{\tau} - \tau, \bar{\eta})|$$
$$\leq \|\hat{\tau} - \tau\| \left(\|\bar{\eta}\| + 1\right) + \|\bar{\gamma}\| \cdot \|\hat{\tau}\| .$$

Similarly we prove that

$$\|\hat{\tau} + \tau\| + \left|\hat{d} + d\right| \leq \|\hat{\tau} + \tau\| \left(\|\bar{\eta}\| + 1\right) + \|\bar{\gamma}\| \cdot \|\hat{\tau}\| .$$

Therefore

$$\min\left\{\|\hat{\tau} - \tau\| + \left|\hat{d} - d\right|, \|\hat{\tau} + \tau\| + \left|\hat{d} + d\right|\right\} \leq \min\{\|\hat{\tau} + \tau\|, \|\hat{\tau} - \tau\|\} \left(\|\bar{\eta}\| + 1\right) + \|\bar{\gamma}\| \cdot \|\hat{\tau}\| . \tag{19}$$

In view of assumptions of Theorem 2, we get

$$\min\{\|\hat{\tau} + \tau\|, \|\hat{\tau} - \tau\|\} \xrightarrow{\mathsf{P}} 0, \quad n \to \infty.$$

Taking into account that $\|\hat{\tau}\| = 1$, $\left\|\bar{\eta}^{(n)}\right\| = O(1)$ as $n \to \infty$, and that $\bar{\gamma}^{(n)} \xrightarrow{\mathsf{P}} 0$, $n \to \infty$, we derive the result of the theorem from (19) by the law of large numbers. □

Remark. The presence of an extra assumption in Theorem 2 saying that the sequence $\bar{\eta}^{(n)}$, $n \geq 1$, is bounded can be explained as follows. Even if the normal vector to the necessary hyperplane is estimated quite precisely, the distance to this hyperplane is not always estimated with the accuracy needed.

4. Sufficient conditions for the strong consistency of the orthogonal regression estimator

Theorem 3. *Assume that conditions* (i) *and* (ii) *hold. We further assume that, for some $r \geq 2$ and $n_0 \geq 1$,*

$$\sup_{i\geq 1} \mathsf{E} \|\gamma_i\|^{2r} < \infty,$$
$$\sum_{n=n_0}^{\infty} \left(\frac{\lambda_{\max}\left(S_{\eta\eta}^{(n)}\right)}{n\lambda_2^2\left(S_{\eta\eta}^{(n)}\right)} \right)^r < \infty,$$
$$\sum_{n=n_0}^{\infty} \left(\frac{1}{\sqrt{n}\lambda_2\left(S_{\eta\eta}^{(n)}\right)} \right)^r < \infty.$$

Then the estimator $\hat{\tau}$ is strictly consistent, that is,

$$\min\{\|\hat{\tau} - \tau\|, \|\hat{\tau} + \tau\|\} \to 0 \qquad a.s.$$

as $n \to \infty$.

Proof. It follows from (11) and (13) that

$$\mathsf{E} \sum_{n=n_0}^{\infty} \|B_n\|^{2r} \cdot \left\|S_{\gamma\eta}^{(n)}\right\|_F^{2r} = \sum_{n=n_0}^{\infty} \mathsf{E} \left\|S_{\gamma\eta}^{(n)}\right\|_F^{2r} \cdot \|B_n\|^{2r} < \infty,$$

whence $\|B_n\| \cdot \left\|S_{\gamma\eta}^{(n)}\right\|_F \to 0$ almost surely as $n \to \infty$. Considering (9) and (10), it remains to prove the convergence $\|B_n\| \cdot \left\|S_{\gamma\gamma}^{(n)} - \Sigma\right\| \to 0$ almost surely as $n \to \infty$. For this, again apply the Rosenthal inequality for $r \ge 2$:

$$\mathsf{E}\left\|\overline{\gamma\gamma^\top}^{(n)} - \Sigma\right\|_F^r = \frac{1}{n^r}\mathsf{E}\left\|\sum_{i=1}^n\left(\gamma_i\gamma_i^\top - \Sigma\right)\right\|_F^r$$
$$\le \frac{\alpha}{n^r}\sum_{i=1}^n \mathsf{E}\left\|\gamma_i\gamma_i^\top - \Sigma\right\|_F^r + \frac{\beta}{n^r}\left(\sum \mathsf{E}\left\|\gamma_i\gamma_i^\top - \Sigma\right\|_F^2\right)^{r/2} = O\left(n^{-r/2}\right).$$

Then (15) and (17) together with assumptions of Theorem 3 imply that

$$\mathsf{E}\sum_{n=n_0}^\infty \|B_n\|^r \|S_{\gamma\gamma} - \Sigma\|_F^r = \sum_{n=n_0}^\infty \mathsf{E}\|S_{\gamma\gamma} - \Sigma\|_F^r \|B_n\|^r < \infty.$$

Therefore $\|B_n\| \cdot \|S_{\gamma\gamma} - \Sigma\|_F \to 0$ almost surely as $n \to \infty$ and this proves the theorem. □

The following result is proved similarly to Theorem 2.

Theorem 4. *Let all assumptions of Theorem* 3 *hold and let the sequence* $\bar\eta^{(n)}$, $n \ge 1$, *be bounded. Then the orthogonal regression estimator* $\left(\hat\tau; \hat d\right)$ *is strictly consistent, that is,*

$$\min\left\{\|\hat\tau - \tau\| + \left|\hat d - d\right|, \|\hat\tau + \tau\| + \left|\hat d + d\right|\right\} \to 0 \quad a.s.$$

as $n \to \infty$.

5. An application to the multiple linear model

The results of the previous sections can be used to study model (1). To transform this model to the form of implicit model (2), we put

$$z_i = \begin{pmatrix} y_i \\ x_i \end{pmatrix}, \qquad \eta_i = \begin{pmatrix} b_0 + x^\top\xi_i \\ \xi_i \end{pmatrix}, \qquad \gamma_i = \begin{pmatrix} \varepsilon_i \\ \delta_i \end{pmatrix}, \qquad i = 1, \dots, n, \tag{20}$$

$$\tau = \left(\frac{1}{\sqrt{1 + \|x\|^2}}; \frac{-x^\top}{\sqrt{1 + \|x\|^2}}\right)^\top, \qquad d = \frac{b_0}{\sqrt{1 + \|x\|^2}}. \tag{21}$$

Then we apply Theorems 1–4 to the transformed model (2) in order to derive some corollaries for model (1). First we describe the corresponding conditions. As far as model (1) is concerned, we keep assumptions (i) and (ii) where γ_i is given by (20).

Using the estimators $\hat\tau$ and $\hat d$ we construct estimators $\hat x$ and $\hat b$ of the regression parameters x and b_0. Let $\hat\tau_1$ denote the first coordinate of the vector $\hat\tau$, while $P\hat\tau$ denotes the remaining part of the vector $\hat\tau$ without the first coordinate. The corresponding notation is introduced for the vector τ, as well. If $\hat\tau_1 \ne 0$, then we set

$$\hat x = \frac{-1}{\hat\tau_1} P\hat\tau, \qquad \hat b = \frac{\hat d}{\hat\tau_1}. \tag{22}$$

Otherwise both estimators are equal to zero.

Now we are ready to state the results.

Theorem 5. *Let assumptions* (i) *and* (ii) *hold for model* (1). *Further, we assume that, for some* $1 \le r \le 2$,

$$\sup_{i\ge1} \mathsf{E}\|\gamma_i\|^{2r} < \infty, \tag{23}$$

$$\frac{\lambda_{\max}(S_{\xi\xi})}{n\lambda_{\min}^2(S_{\xi\xi})} \to 0, \qquad n \to \infty, \tag{24}$$

$$n^{1-1/r}\lambda_{\min}(S_{\xi\xi}) \to \infty, \qquad n \to \infty. \tag{25}$$

Then $\hat{x} \xrightarrow{\mathsf{P}} x$ *as* $n \to \infty$.

Theorem 6. *Assume that all assumptions of Theorem* 5 *hold and that the sequence* $\bar{\xi}^{(n)}$, $n \geq 1$, *is bounded. Then* $\hat{b} \xrightarrow{\mathsf{P}} b_0$ *as* $n \to \infty$.

Theorem 7. *Let conditions* (i) *and* (ii) *hold for model* (1). *We further assume that, for some* $r \geq 2$ *and* $n_0 \geq 1$,

$$\sup_{i\geq 1} \mathsf{E}\,\|\gamma_i\|^{2r} < \infty,$$

$$\sum_{n=n_0}^{\infty} \left(\frac{\lambda_{\max}(S_{\xi\xi})}{n\lambda_{\min}^2(S_{\xi\xi})}\right)^r < \infty, \tag{26}$$

$$\sum_{n=n_0}^{\infty} \left(\frac{1}{\sqrt{n}\lambda_{\min}(S_{\xi\xi})}\right)^r < \infty. \tag{27}$$

Then $\hat{x} \to x$ *almost surely as* $n \to \infty$.

Theorem 8. *Let assumptions of Theorem* 7 *hold and let the sequence* $\bar{\xi}^{(n)}$, $n \geq 1$, *be bounded. Then* $\hat{b} \to b_0$ *almost surely as* $n \to \infty$.

To prove Theorems 5–8, we first discuss the assumptions imposed on the eigenvalues of the matrix $S_{\xi\xi}$. Since

$$S_{\eta\eta} = \begin{pmatrix} x^\top S_{\xi\xi} x & x^\top S_{\xi\xi} \\ S_{\xi\xi} x & S_{\xi\xi} \end{pmatrix} = \begin{pmatrix} 1 & x^\top \\ 0 & I_{m-1} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & S_{\xi\xi} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ x & I_{m-1} \end{pmatrix},$$

the Courant–Fisher principle of the minimax characterization of eigenvalues of a symmetric matrix (see [4]) implies that $\lambda_{\min}(S_{\xi\xi}) \leq \text{const} \cdot \lambda_2(S_{\eta\eta})$. Further,

$$\lambda_{\max}(S_{\eta\eta}) \leq \operatorname{tr}(S_{\eta\eta}) \leq \left(1 + \|x\|^2\right) \operatorname{tr}(S_{\xi\xi}) \leq m \left(1 + \|x\|^2\right) \lambda_{\max}(S_{\xi\xi}).$$

This proves that if assumptions of one of Theorems 5–8 are satisfied, then so are the assumptions of the corresponding theorem among Theorems 1–4 for the transformed model (2).

Considering the estimators in (22), we see that if the estimators $\hat{\tau}$ and $\hat{d}$ are changed by the opposite ones, then the expressions for $\hat{x}$ and $\hat{b}$ are not changed. Since

$$x = \frac{-P\tau}{\tau_1}, \qquad b_0 = \frac{d}{\tau_1}, \qquad \tau_1 \neq 0,$$

the convergence $\min\{\|\hat{\tau} - \tau\|, \|\hat{\tau} + \tau\|\} \to 0$ as $n \to \infty$ (either in probability or almost surely) implies the corresponding convergence $\hat{x} \to x$ as $n \to \infty$. Similarly we show that the result of every of the Theorems 1–4 for implicit model (2) implies the result of the corresponding theorem among Theorems 5–8 for model (1).

Therefore, Theorems 5–8 are corollaries of Theorems 1–4, indeed.

6. A comparison with known results on consistency

Our current aim is to compare Theorems 5–8 with known results on the consistency of total least squares estimators in the case where our theorems are applied for the transformed regression model without free term

$$\begin{cases} b_i = x_0^\top a_i^0 + \tilde{b}_i, \\ a_i = a_i^0 + \tilde{a}_i, \end{cases} \tag{28}$$

which is obtained from model (1) by using the substitutions

$$a_i = \begin{pmatrix} 1 \\ x_i \end{pmatrix}, \qquad a_i^0 = \begin{pmatrix} 1 \\ \xi_i \end{pmatrix}, \qquad \tilde{a}_i = \begin{pmatrix} 0 \\ \delta_i \end{pmatrix}, \qquad b_i = y_i, \qquad x_0 = \begin{pmatrix} b_0 \\ x \end{pmatrix}, \tag{29}$$
$$\tilde{b}_i = \varepsilon_i, \qquad i = 1, \dots, n.$$

Put $\tilde{c}_i = \left(\tilde{a}_i^\top; \tilde{b}_i\right)^\top$. Then assumptions (i) and (ii) concerning model (1) are rewritten as follows in the case under consideration:

(iii) The vectors $\tilde{c}_i$, $i \geq 1$, are jointly independent and have zero mean.
(iv) The vectors $\tilde{c}_i$, $i \geq 1$, have an identical covariance matrix

$$\mathbf{cov}(\tilde{c}_i) = \sigma^2 \cdot \operatorname{diag}(0, 1, 1, \dots, 1)$$

where $\sigma^2 > 0$ is unknown.

The total least squares estimator $\hat{x}_{\mathrm{TLS}}$ of the parameter x_0 in model (28) is a solution of the optimization problem

$$\min_{(\Delta a_i \in \mathbb{R}^m,\, \Delta b_i \in \mathbb{R})} \sum_{i=1}^{n} \left(\|\Delta a_i\|^2 + |\Delta b_i|^2 \right)$$

if there exists a vector $x \in \mathbb{R}^m$ such that

$$b_i - \Delta b_i = x^\top (a_i - \Delta a_i), \qquad i = 1, \dots, n.$$

The least restrictive sufficient conditions for the consistency as well as for the strong consistency of the total least squares estimator for model (28) are given in the paper [7] under assumptions (iii) and (iv). For convenience, we provide the statements of the corresponding two theorems from [7] with $A_0 := [a_1^0, a_2^0, \dots, a_n^0]^\top$.

Theorem 9. *Let conditions* (iii) *and* (iv) *hold for model* (28). *Further assume that, for some* $1 \leq r \leq 2$,

$$\sup_{i \geq 1} \mathsf{E} \|\tilde{c}_i\|^{2r} < \infty, \tag{30}$$

$$n^{-1/r} \lambda_{\min}\left(A_0^\top A_0\right) \to \infty, \qquad n \to \infty. \tag{31}$$

Then $\hat{x}_{\mathrm{TLS}} \xrightarrow{\mathsf{P}} x_0$ *as* $n \to \infty$.

Theorem 10. *Let conditions* (iii) *and* (iv) *hold for model* (28). *Further assume that, for some* $r \geq 2$ *and* $n_0 \geq 1$,

$$\sup_{i \geq 1} \mathsf{E} \|\tilde{c}_i\|^{2r} < \infty,$$
$$\sum_{n=n_0}^{\infty} \left(\frac{\sqrt{n}}{\lambda_{\min}(A_0^\top A_0)} \right)^r < \infty. \tag{32}$$

Then $\hat{x}_{\mathrm{TLS}} \to x_0$ *almost surely as* $n \to \infty$.

Since Theorems 9 and 10 provide the consistency for estimators of both parameters b_0 and x, we compare these results with our Theorems 6 and 8. Note that assumptions (23) and (30) are the same, while conditions (31) and (32) follow from (25) and (27), respectively, if $\lambda_{\min}(S_{\xi\xi})$ is changed by $\frac{1}{n}\lambda_{\min}\left(A_0^\top A_0\right)$. First we consider the matrix $A_0^\top A_0$. In view of notation (29), we get

$$A_0 = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \xi_1 & \xi_2 & \dots & \xi_n \end{pmatrix}^\top.$$

Then

$$\frac{1}{n}A_0^\top A_0 = \frac{1}{n}\sum_{i=1}^n a_i^0 {a_i^0}^\top = \begin{pmatrix} 1 & \bar{\xi}^\top \\ \bar{\xi} & \overline{\xi\xi^\top} \end{pmatrix}.$$

To compare the numbers $\lambda_{\min}(S_{\xi\xi})$ and $\frac{1}{n}\lambda_{\min}\left(A_0^\top A_0\right)$ we use properties of symmetric non-negative definite matrices. For such a matrix $A \in \mathbb{R}^{m\times m}$, denote

$$\det(A-\lambda I_m) = (-1)^m\lambda^m + \cdots + b(A)\lambda^2 - c(A)\lambda + \det(A). \tag{33}$$

Then $c(A)$ is equal to the sum of the cofactors of the diagonal entries of the matrix A. Now we order the eigenvalues of the matrix A in the ascending order:

$$0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_m. \tag{34}$$

By Vieta's theorem

$$\det(A) = \lambda_1\lambda_2\dots\lambda_m, \qquad c(A) = \sum_{j=1}^m \frac{\lambda_1\lambda_2\dots\lambda_m}{\lambda_j}. \tag{35}$$

Since the eigenvalues are non-decreasing,

$$\frac{1}{m}\lambda_1 = \frac{\lambda_1\lambda_2\dots\lambda_m}{m\lambda_2\lambda_3\dots\lambda_m} \le \frac{\det(A)}{c(A)} \le \frac{\lambda_1\lambda_2\dots\lambda_m}{\lambda_2\lambda_3\dots\lambda_m} = \lambda_1. \tag{36}$$

If, for every $i=\overline{2,m}$, we subtract the first row of the matrix A multiplied by the $(i-1)$th element of the vector $\bar{\xi}$ from the ith row of the matrix $\frac{1}{n}A_0^\top A_0$, then we conclude that

$$\det\left(\frac{1}{n}A_0^\top A_0\right) = \det\left(S_{\xi\xi}\right). \tag{37}$$

Using the latter result for the dimension $m-1$ we obtain

$$c\left(\frac{1}{n}A_0^\top A_0\right) = c(S_{\xi\xi}) + \det\left(\overline{\xi\xi^\top}\right) \ge c(S_{\xi\xi}). \tag{38}$$

Taking into account (37) and (38) we derive from (36) that

$$\lambda_{\min}\left(\frac{1}{n}A_0^\top A_0\right) \le \frac{m\det\left(\frac{1}{n}A_0^\top A_0\right)}{c\left(\frac{1}{n}A_0^\top A_0\right)} \le \frac{m\det(S_{\xi\xi})}{c(S_{\xi\xi})} \le m\lambda_{\min}(S_{\xi\xi}). \tag{39}$$

This means that conditions (25) and (27) are less restrictive than conditions (31) and (32), respectively. For the typical case where $\lambda_{\max}(S_{\xi\xi}) = O(\lambda_{\min}(S_{\xi\xi}))$ as $n\to\infty$, conditions (24) and (26) follow from conditions (25) and (27), respectively.

7. Concluding remarks

We considered a linear regression implicit model with errors in variables and found comparatively mild conditions for the consistency of the orthogonal regression estimator. As a corollary, we obtained results concerning the consistency of the total least squares estimator in the multiple linear regression with a scalar response, free term, and errors in variables. Note that these results do not follow from the corresponding results of [7]. It is of interest to investigate the total least squares estimator further in the corresponding multiple regression with a vector response and free term.

Bibliography

[1] W. Fuller, *Measurement Error Models*, Wiley, New York, 1987. MR898653

[2] P. Gallo, *Consistency of regression estimates when some variables are subject to errors*, Comm. Statist. Theory Methods **11** (1982), no. 9, 973–983. MR655466

[3] L. J. Gleser, *Estimation in multivariate "errors-in-variables" regression model: large sample results*, Ann. Statist. **9** (1981), no. 1, 24–44. MR600530

[4] A. I. Kostrikin and Yu. I. Manin, *Linear Algebra and Geometry*, Series in Algebra, Logic and Applications, vol. 1, CRC Press, 1989. MR1057342

[5] A. Kukush and S. Van Huffel, *Consistency of elementwise-weighted total lest squares estimator in a mulivariate errors-in-variables model* $AX = B$, Metrika **59** (2004), no. 1, 75–97. MR2043433

[6] S. V. Masiuk, A. G. Kukush, S. V. Shklyar, M. I. Chepurny, I. A. Likhtarov, *Radiation Risk Estimation: Based on measurement error models*, 2nd ed., De Gruyter Ser. Math. Life Sci., vol. 5, De Gruyter, 2017. MR3726857

[7] S. V. Shklyar, *Conditions for the consistency of the total least squares estimator in an errors-in-variables linear regression model*, Theory Probab. Math. Statist. **83** (2011), 175–190. MR2768857

Department of Mathematical Analysis, Faculty for Mechanics and Mathematics, Kyiv Taras Shevchenko National University, Volodymyrs'ka Street, 64/13, Kyiv 01601, Ukraine
Email address: oodashkov@gmail.com

Department of Mathematical Analysis, Faculty for Mechanics and Mathematics, Kyiv Taras Shevchenko National University, Volodymyrs'ka Street, 64/13, Kyiv 01601, Ukraine
Email address: alexander_kukush@univ.kiev.ua

Received 04/OCT/2017
Translated by S. V. KVASKO