University of Economics, Prague
Faculty of Informatics and Statistics
Department of Econometrics

Technical Report 04/16

Linear regression with bounded errors in data: Total 'least squares' with the Chebyshev norm

Milan Hladík, Michal Černý and Jaromír Antoch

Preprint

October 2016

Statistical Papers manuscript No.
(will be inserted by the editor)

EIV regression with bounded errors in data: Total 'Least Squares' with Chebyshev norm

Milan Hladík · Michal Černý · Jaromír Antoch

Received: date / Accepted: date

Abstract We consider the linear regression model with stochastic regressors and stochastic errors both in regressors and the dependent variable ("structural EIV model"), where the regressors and errors are assumed to satisfy some interesting and general conditions, different from traditional assumptions on EIV models (such as Deming regression). The most interesting fact is that we need neither independence of errors, nor identical distributions, nor zero means. The first main result is that the TLS estimator, where the traditional Frobenius norm is replaced by the Chebyshev norm, yields a consistent estimator of regression parameters under the assumptions summarized below. The second main result is that we design an algorithm for computation of the estimator, reducing the computation to a family of generalized linear-fractional programming problems (which are easily computable by interior point methods). The conditions under which our estimator works are (said rougly): it is known which regressors are affected by random errors and which are observed exactly; that the regressors satisfy a certain asymptotic regularity condition; all error distributions, both in regressors and in the endogenous variable, are bounded in absolute value by a common bound (but the bound is unknown and is estimated); there is a high probability that we observe a family of data points where the errors are close to the bound. We also generalize the method

M. Hladík
Department of Applied Mathematics, Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic, Malostranské náměstí 25, CZ11800 Prague
E-mail: milan.hladik@matfyz.cz

M. Černý
Department of Econometrics, Faculty of Computer Science and Statistics, University of Economics in Prague, Czech Republic, Winston Churchill Square 4, CZ13067 Prague
E-mail: cernym@vse.cz

J. Antoch
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Czech Republic, Sokolovská 83, CZ18675 Prague
E-mail: antoch@karlin.mff.cuni.cz

to the case that the bounds of errors in the dependent variable and regressors are not the same, but their ratios are known or estimable. The assumptions, under which our estimator works, cover many settings where the traditional TLS is inconsistent.

Keywords Errors-In-Variables · Measurement error models · Total Least Squares · Chebyshev matrix norm · Bounded error distributions · Generalized Linear-Fractional Programming

Mathematics Subject Classification (2000) 62J05 · 65C60 · 90C32

1 Introduction

It is generally known that if a linear regression model suffers from the errors-in-variables (EIV) problem (that is, if we can observe only a contaminated form of regressors), then ordinary least squares (OLS) is inconsistent. Various methods for EIV estimation have been studied in econometric literature, including the instrumental variables (IV) method [8,9,27], or the generalized method of moments [3,7,9,10,12]. We also refer to the recent review article [5] and references therein, including many applications of EIV models in finance, with the prominent example of the capital asset pricing model. See also [1,2, 19,23,28] for further applications and special problems such as EIV in dynamic or panel models. Recall also that further important problems are related to EIV regression models, such as identifiability issues or, more generally, the necessity of additional information for estimation procedures (such as parameter restrictions, moment restrictions or additional data as instruments). For example, see [36,37] for the linear case where identifiability is assured via prior restrictions on parameters, and see also a remarkable result [22] on logistic regression. Our setup is essentially different from the linear models studied in [36,37] (details of our model are given is Section 3). But still, identifiability issues arise in our case as well, as discussed in Section 8 of this article.

The motivation for the construction of our estimator can be also illustrated by a quote from the introduction of [12] summarizing the 'standard' treatment of the EIV problem in econometrics:

> "The most common remedy is to use economic theory or intuition to find additional observable variables that can serve as instruments, but in many situations no such variables are available. Consistent estimators based on the original, unaugmented set of observable variables are therefore potentially quite valuable."

Our estimator fits in this context: it needs only the unaugmented set of observable variables.

In the terminology of [5], this paper is a contribution to the theory of "classical estimators". More precisely, this is a contribution to the area of minimum-norm estimators. From now on, we will restrict our attention only to this class of estimators. The introduction will be devoted to a discussion on the usage of various norms, showing how this area is complemented by our approach.

Remark. Now it is possible to skip the rest of the introduction and continue in Section 3, where model assumptions are stated formally. In Section 3, only equations (1) – (3) are needed.

Notation. We consider the linear regression model

$$y_i = x_i'\beta + \varepsilon_i, \quad i = 1, \dots, n, \tag{1}$$

where $\beta \in \mathbb{R}^p$ is an unknown vector of parameters, $x_i = \big(x_{i1}, \dots, x_{ip}\big)'$ are unobservable stochastic regressors, and $\varepsilon_i,\ i = 1, \dots, n$, are additive random errors (in observations of the dependent variable). We assume the following form of the EIV model: observable data are $\big(y_i, z_{ij}\big)$, where

$$z_{ij} = x_{ij} + \nu_{ij}, \quad i = 1, \dots, n,\ j = 1, \dots p, \tag{2}$$

and ν_{ij} are random errors (in observations of the regressors). Section 3 will be devoted to a detailed formulation of assumptions on the distribution of $(x_{ij}, \nu_{ij}, \varepsilon_i)$.

We denote below

$$Z^n = \big(z_{ij}\big)_{i=1,\dots,n}^{j=1,\dots p} \qquad \text{and} \qquad y^n = (y_1, \dots, y_n)'. \tag{3}$$

An overview: EIV regression models and minimum-norm estimators. The EIV models, also called *measurement error models*, have a very long history, with Addock (1877, 1878) usually being regarded as the first person to specifically consider them. The concepts have been independently developed for solving different problems arising especially in statistics and numerical mathematics. Corresponding methods, especially the so-called total least squares approach and its modifications, became quite popular in the 1980s after the influential paper published by [17]. Interest was especially high due to the fact that computationally stable and efficient methods based on SVD decomposition had been developed. In the 1990s a number of extensions were suggested, especially by the groups around Cheng and Van Huffel.

Total Least Squares (TLS). Recall that the *total least squares (TLS) problem* can be algebraically formulated as follows. Given $A \in \mathbb{R}^{n\times p}$ and $w \in \mathbb{R}^n$, find $\Delta A \in \mathbb{R}^{n\times p}$ and $\Delta w \in \mathbb{R}^n$ such that:

- the linear system $(A + \Delta A)\xi = (w + \Delta w)$ is solvable;
- $\|(\Delta A, \Delta w)\|_F$ is minimal, where $\|\cdot\|_F$ denotes the Frobenius norm.

It is worth noting that the literature about EIV, TLS and their generalizations focuses mostly on estimation and numerical algorithms, and less on statistical properties such as consistency, asymptotic distribution, testing, developing confidence intervals, etc. Most of the results, some of which can now be considered classical, can be found in [4,6,39], among others.

Assume the EIV model in which rows $(\varepsilon_i, \nu_{i1}, \dots, \nu_{ip})$ of the matrix of errors are independent, while not necessarily homoscedastic, and the TLS approach based on the Frobenius norm leading to the estimate $\widehat{\beta}$ of β. Then existence of $\widehat{\beta}$ with the probability tending to one as $n \to \infty$ was proved by [16,31].

Asymptotic properties of $\widehat{\beta}$ have been studied by many authors. The maximum likelihood approach under normality assumption on errors was developed by Healy [20]. However, it appears that the first really deep study of consistency and asymptotic distribution can be credited to Gleser [16]. He has shown that when $\mathrm{var}(\varepsilon) = \sigma^2 I$, a wide class of approaches based on least squares in model (1) for estimation of the vector of parameters lead to identical estimates and that they coincide with the maximum likelihood one under the assumption of normality of errors. Moreover, he has established (not assuming normality of errors) quite general conditions under which $\widehat{\beta}$ in the EIV model with or without intercept are strongly consistent and asymptotically normal. Recall that, in his approach, one of the key conditions is that the matrix $\Xi = \lim_{n\to\infty} n^{-1} Z^{n\prime} Z^n$ exists and is positive definitive.

It appears that the conditions under which $\widehat{\beta}$ is strongly consistent are too restrictive. Thus, Gallo [15] has derived weaker conditions under which $\widehat{\beta}$ is weakly consistent. More precisely, he has proven that, provided rows of the matrix of errors $[\,\varepsilon_i, \nu_{i1}, \ldots, \nu_{ip}\,]$ are independent (row by row), their distribution has finite fourth moments and

$$\frac{1}{\sqrt{n}} \lambda_{min}\left(Z^{n\prime} Z^n\right) \to \infty \quad \& \quad \frac{\lambda_{min}^2\left(Z^{n\prime} Z^n\right)}{\lambda_{max}\left(Z^{n\prime} Z^n\right)} \to \infty \quad \text{as} \quad n \to \infty,$$

where $\lambda_{min}(A)$ $\big(\lambda_{max}(A)\big)$ is the smallest (largest) eigenvalue of matrix A, then $\widehat{\beta}$ is weakly consistent, i.e., $\widehat{\beta} \xrightarrow{P} \beta$ as $n \to \infty$. He has also observed that OLS in the EIV model is inconsistent and a TLS estimate should be used instead.

The results for weak and strong consistency were later strengthened by Kukush [24]. Their conditions are on the one hand weaker, yet on the other hand not so easily verifiable.

As mentioned above, the first asymptotic results for the TLS estimates were obtained under quite strict and restrictive conditions, one of them being that the errors are independent and identically distributed. Therefore, generalized TLS (GTLS) have been suggested, and their properties studied, in which this assumption is relaxed and replaced by the condition that the errors are row-wise independent but correlated within the rows with identical covariance matrix V. It has been shown that this situation can be reduced to the TLS problem by multiplying the data matrix by $V^{-1/2}$. However, even the conditions of GTLS method are too restrictive for some applications, especially due to the assumption of the equal covariance of all rows in the error matrix. Therefore, further generalizations have been suggested in the literature. Among them we would like to emphasize a proposal termed EW-TLS, by Kukush and Van Huffel [25], who assumed that the errors are row-wise correlated with known correlation matrices. Evidently, any generalization of this type has to be paid for by much more complicated numerical calculations, because for the EW-TLS no closed-form solution exists, and one has to apply non-convex optimization methods when calculating its approximation. Later on Markowski et al. [29] went even further, allowing that correlations among the errors within each row exist with possibly singular covariance matrices. Recently, Pešta [33] has established conditions under which the TLS estimate is strongly consistent under the assumption that the errors are weakly dependent and form a so-called α- or φ-mixing sequence.

Reformulation of TLS with other matrix norms. The TLS estimates are usually based on the Frobenius norm. However, other matrix norms can be used as well. Among them, an important role is played by the class of orthogonally invariant (unitarily invariant) norms[1]. Recall that these norms were already completely characterized by von Neumann [40]. We also refer to the recent article [34] on unitarily invariant norms in EIV estimation.

Let $A \in \mathbb{R}^{n\times p}$ be any matrix with singular values $\sigma_1 \geq \ldots \geq \sigma_{\min\{n,p\}} \geq 0$. Among the most popular representatives of the orthogonally invariant norms belong *q-Ky Fan k-norms*, defined as

$$\|A\|_q^{(k)} = \left(\sum\nolimits_{i=1}^{k} \sigma_i^q\right)^{1/q}, \quad q \geq 1,\ 1 \leq k \leq \min\{n,p\}$$

with the most important representative being the spectral matrix norm ($k = 1$ and $q = 1$), Frobenius norm ($k = \min\{n,p\}$ and $q = 2$) or nuclear matrix norm ($k = \min\{n,p\}$ and $q = 1$). It is evident that popular q-Schatten norms can also be considered a special case of q-Ky Fan k-norms for $k = \min\{n,p\}$. For details see e.g. [35].

It follows from the results of [16, 31] that when the distribution of the errors is absolutely continuous with respect to the Lebesque measure and any unitarily invariant norm is used instead of the Frobenius norm in the EIV model, $\widehat{\beta}$ exists with probability equal to one as $n \to \infty$ and coincides with the classical TLS estimate. Moreover, asymptotic properties of the classical EIV models hold, or rather, can be straightforwardly generalized to this situation as well.

2 Our contribution and organization of the paper

The central role of this paper will be played by the Chebyshev matrix norm $\|A\|_{\max} = \max_{i,j} |A_{ij}|$. This paper complements the above mentioned theory, since the Chebyshev norm does not fit in the previous context and similar theory is not currently available for this norm (as far as we are aware). We will show that the Chebyshev norm can be used for consistent estimation of the regression parameters of the EIV models under conditions very different from those discussed above. The conditions will be summarized in Section 3. We can motivate them by the following easy example.

An example: a setup where our estimator is consistent while OLS is not. Consider the simple model

$$y_i = \beta + \varepsilon_i, \quad i = 1, \ldots, n.$$

[1] A matrix norm $\|\cdot\|$ is orthogonally invariant, if $\|UAV\| = \|A\|$ for all $A \in \mathbb{R}^{n\times p}$ and all unitary matrices $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{p\times p}$.

Assume that ε_i are independent, identically and continuously distributed with support $(-\gamma, \gamma)$, where γ is an unknown constant. Assume further that $\mathsf{E}\varepsilon_i \neq 0$. In this setup, we have no errors in regressors, and thus TLS reduces to OLS. And OLS is obviously inconsistent. But our method yields a consistent estimator of β and γ.

Outline of the paper. This paper is organized as follows. First, in Section 3 conditions on the design matrix and errors are formulated. In Section 4 the original TLS problem is reformulated with the Chebyshev norm used instead of the Frobenius one. The main results are stated in Section 5. First, we show that if we can solve the Chebyshev Norm Problem modification (CNP), then we get consistent estimators of the regression coefficients. Second, we show how to compute the desired estimates. Section 6 is devoted to a discussion about computational complexity of the proposed algorithm. Proofs of theorems can be found in Section 7. General discussion and conclusions follow.

3 Model assumptions

Now we formulate two assumptions on the distribution of $(x_{ij}, \nu_{ij}, \varepsilon_i)$, under which we will derive a consistent estimator for β based on total least squares, where the usual Frobenius norm is replaced by the Chebyshev norm. Then we propose an algorithm for its computation.

Assumption 1. There exists an unknown constant $\gamma > 0$, called *radius*, such that

(i) $|\varepsilon_i| \leq \gamma$ a.s., $i = 1, \dots, n$;

and there is a known index set $\mathit{\Gamma} \subseteq \{1, \dots, p\}$ such that for all $j = 1, \dots, p$:

(ii) if $j \notin \mathit{\Gamma}$, then $\nu_{ij} = 0$ a.s. for all $i = 1, \dots, n$;
(iii) if $j \in \mathit{\Gamma}$, then $|\nu_{ij}| \leq \gamma$ a.s. for all $i = 1, \dots, n$.

Comment. The set $\mathit{\Gamma}$ formalizes the assumption that we know in advance which regressors are measured exactly and which are not.

EIV models in which some explanatory variables are subject to errors and some are measured exactly appear quite often and naturally in practice. They bring about not at all trivial problems for the theory when $\mathit{\Gamma}$ is assumed to be known. (Conversely, if we have only partial information about $\mathit{\Gamma}$, or it is completely unknown, the situation would be much more complicated and this case would deserve further research.)

The simplest situation when some regressors are known to be exact arises when we include a non-random (fixed) intercept into the model. In the literature, these models are sometimes termed PEIV (partial errors-in-variables) models and have been studied by several authors. From the numerical and algorithmic points of view the PEIV model was considered e.g., by [18], who obtained a so-called LS-TLS estimate. The basic idea of their approach was to separate exact and approximate observations from each other, and to solve the resulting rank-deficient optimization problem. A similar situation was regarded from the statistical point of view already by [16], who studied the EIV model both with and without the intercept. It is important to point out that properties of the classical EIV models can be straightforwardly generalized to the PEIV models as well.

Assumption 2 (asymptotic properties of regressors and errors). Let $\|\cdot\|$ be any vector norm. We assume that

$$(\forall \alpha > 0)\ (\exists c > 0)\ (\forall u \in \mathbb{R}^p \text{ s.t. } \|u\| = 1)\ \Pr[\mathcal{A}_n(\alpha, c, u)] \stackrel{n\to\infty}{\longrightarrow} 1,$$

where $\mathcal{A}_n(\alpha, c, u)$ is the following event: there exists $i_0 \in \{1, \ldots, n\}$ such that

(i) $|x'_{i_0} u| \geq c$; and
(ii) $-\mathrm{sgn}(x'_{i_0} u) \cdot \varepsilon_{i_0} \geq \gamma - \alpha$; and
(iii) $(\forall j \in \Gamma)\ \mathrm{sgn}(x'_{i_0} u) \cdot \mathrm{sgn}(\beta_j + u_j) \cdot \nu_{i_0 j} \geq \gamma - \alpha$,

where $\mathrm{sgn}(\xi) = 1$ if $\xi \geq 0$ and $\mathrm{sgn}(\xi) = -1$ if $\xi < 0$.

Remark. A generalization of Assumptions 1 and 2 for the case of non-identical error radii will be considered in Section 8.

Informal discussion on Assumption 2. Item 2(i) postulates a kind of asymptotic regularity conditions for regressors. Items 2(ii-iii) are the crucial ones. We have stated them in the weakest possible form necessary for the forthcoming proofs. However, roughly stated, (ii-iii) require that if n is sufficiently large, then we have a high probability that there appears an observation i_0 where the errors approach the limits $\pm\gamma$ arbitrarily close. Or, said in another way, whichever signs ± 1 of errors we prescribe, we have a high probability that we will meet an observation i_0 where the errors are arbitrarily close to the limits $\pm\gamma$ in the prescribed directions.

Observe that we need neither independence, nor zero means, nor identical distributions.

Example 1. When all error terms ε_i and ν_{ij} with $j \in \Gamma$ have continuous distributions with support $(-\gamma, \gamma)$ (but they need not be identically distributed) and are independent, then Assumptions 2(ii-iii) are satisfied.

Example 2. The distributions of ε_i, ν_{ij} with $j \in \Gamma$ can also be discrete or mixed; for example, if the values $\pm\gamma$ are attained with probability tending to one as $n \to \infty$, then Assumptions 2(ii-iii) are satisfied.

Example 3. We show a "pathological" example when Assumption 2 is not satisfied. Consider the case $p = 1$, $\Gamma = \{1\}$, $\beta_1 > 0$. Then only two choices are possible: $u = 1$ and $u = -1$. If the errors ν_{i1} and the regressors x_i are dependent, almost surely satisfying ($\nu_{i1} \geq 0$ iff $x_i < 0$) and ($\nu_{i1} < 0$ iff $x_i \geq 0$), then Assumption 2 is violated.

Example 3 shows that rather strong and unnatural dependence structures between the regressors and the error terms result in a violation of the assumption.

4 The Chebyshev Norm Problem: Formulation

As already outlined, we will need a reformulation of the original TLS problem when the Frobenius norm is replaced by the Chebyshev norm. We will take the advantage of the following algebraic formulation:

Chebyshev Norm Problem (CNP). Given $A \in \mathbb{R}^{n\times p}$, $w \in \mathbb{R}^n$, and $\mathit{\Gamma} \subseteq \{1,\dots,p\}$, find $\mathit{\Delta} A \in \mathbb{R}^{n\times p}$ and $\mathit{\Delta} w \in \mathbb{R}^n$ such that:

- the linear system $(A + \mathit{\Delta} A)\xi = (w + \mathit{\Delta} w)$ is solvable;
- if $j \notin \mathit{\Gamma}$, then the j-th column of $\mathit{\Delta} A$ is zero; and
- $\|(\mathit{\Delta} A, \mathit{\Delta} w)\|_{\max}$ is minimal.

Notation: definition of $\xi^*(A, w, \mathit{\Gamma})$, $\delta^*(A, w, \mathit{\Gamma})$. When $(\mathit{\Delta} A, \mathit{\Delta} w)$ is (any) solution to CNP with data $(A, w, \mathit{\Gamma})$, then

- $\xi^*(A, w, \mathit{\Gamma})$ denotes (any) solution ξ of the system $(A+\mathit{\Delta} A)\xi = (w+\mathit{\Delta} w)$,
- $\delta^*(A, w, \mathit{\Gamma}) = \|(\mathit{\Delta} A, \mathit{\Delta} w)\|_{\max}$.

5 Main results

The first main result of this paper shows that if we can solve CNP, then we get consistent estimators of β and γ (recall that γ is the radius of the error distribution defined in Assumption 1). The second main result is algorithmic: it will show how to compute the estimators. We continue with the notation of Sections 3 and 4; recalling (3), $Z^n = (z_{ij})$ are the observable values of regressors, which are "contaminated" by errors if $j \in \mathit{\Gamma}$, and are observed precisely if $j \notin \mathit{\Gamma}$.

Theorem 1 *Let*

$$\widehat{\beta}^n = \xi^*(Z^n, y^n, \mathit{\Gamma}) \quad \text{and} \quad \widehat{\gamma}^n = \delta^*(Z^n, y^n, \mathit{\Gamma}). \tag{4}$$

Under Assumptions 1 and 2,

$$\widehat{\beta}^n \xrightarrow{P} \beta \qquad \text{and} \qquad \widehat{\gamma}^n \xrightarrow{P} \gamma \qquad \text{as} \qquad n \to \infty.$$

□

The algorithm for computation of $\widehat{\beta}^n$ and $\widehat{\gamma}^n$ is given by the following result.[2]

Theorem 2 *Let $z_1', \dots, z_n'$ be the rows of Z^n. For a sign vector $s \in \{\pm 1\}^p$, consider the optimization problem*

$$c_s^n = \min_{b\in\mathbb{R}^p} \left\{ \max_{\substack{i\in\{1,\dots,n\}\\ k\in\{0,1\}}} \frac{(-1)^{1-k} z_i' b + (-1)^k y_i}{\eta' D_s b + 1} \,\middle|\, D_s b \geq 0 \right\}, \tag{5}$$

[2] A preliminary version of the algorithm for CNP, presented in Theorem 2, was reported at the ICCS'15 conference and is reported in the proceedings [21]. It was studied there from the complexity-theoretic perspective. Here we present it with a proof for the simple reason that the geometry on which the algorithm is based will also be necessary for the proof of Theorem 1 and cannot be avoided. This paper is a follow-up of the mentioned conference contribution.

where $D_s = \mathrm{diag}(s)$ *and* $\eta = (\eta_1, \ldots, \eta_p)'$ *with*

$$\eta_i = \begin{cases} 1 \text{ if } i \in \Gamma, \\ 0 \text{ if } i \notin \Gamma. \end{cases} \tag{6}$$

Let b_s^n *denote the corresponding argmin (i.e., any* b *satisfying* $D_s b \geq 0$ *for which the minimum value* c_s^n *is attained). Then*

$$\widehat{\gamma}^n = \min_{s \in \{\pm 1\}^p} c_s^n, \tag{7}$$

and if s^* *is the argmin of (7), then* $\widehat{\beta}^n = b_{s^*}^n$*.* □

6 Discussion about computational complexity

Observe that the optimization problem (5) is a generalized linear-fractional programming (GLFP) problem, which can be solved efficiently (in polynomial time) by interior point methods [13,30]. Thus the computation is reduced to solving 2^p GLFPs.

The main question is whether the resulting computation time

$$2^p \times \text{(polynomial computation time for GLFP)}$$

is "good news" or "bad news". Although it is an exponential-time algorithm, we give a complexity-theoretic argument that this is the best that can be achieved. In [21] we proved that solving CNP is an NP-hard problem, and thus not only the algorithm of Theorem 2, but *any* algorithm for CNP must be somehow exponential (unless P = NP). In principle, it can be exponential either in n, the number of observations, or in p, the number of parameters. Since usually $p \ll n$ in practice, the fact that the time is exponential in p and not in n *should be understood as good news.* In practice we usually encounter regression models with a bounded number of regressors (say, at most 20); hence we are to solve at most 2^{20} GLFPs. And 2^{20} is large, but still tractable with the aid of contemporary hardware.

Moreover, observe that the method is easy-to-parallelize, which makes the method suitable for distributed architectures, which are developing very fast nowadays. Indeed, if we can use 2^p parallel processors, each of them is assigned an $s \in \{\pm 1\}^p$, and the processors can solve the 2^p GLFPs (5) independently.

In theory, when $n \to \infty$, the complexity depends on the relationship between n and p. In Section 3 we declared p as a fixed constant; then 2^p is also a fixed constant and the algorithm is polynomial. If we admit that p is a function of n, then it remains polynomial as long as $p = O(\log n)$.

6.1 An important special case: the signs of β are known a priori

There is a special case when the algorithm is even more efficient: it is the case when we know a priori the signs of the regression coefficients. If, for example, we know a priori that $\beta \geq 0$, then it suffices to use Theorem 2 with a single choice $s = (1, \ldots, 1)$ instead of all choices $s \in \{\pm 1\}^p$. Then the problem is reduced to a single GLFP. Thus:

Corollary 1 *If the signs of regression parameters β are known, then the estimators $\widehat{\beta}^n$ and $\widehat{\gamma}^n$ are computable in polynomial time.* □

7 Proofs of Theorems 1 and 2

7.1 Conventions and notation

Let e denote the all-one vector. For a matrix A, its (i, j)-th element is denoted A_{ij}. An inequality $A \leq B$ between matrices is understood entrywise (i.e., $A_{ij} \leq B_{ij}$ for all i, j). The absolute value $|x|$ of a vector x is also understood entrywise. Observe that for every vector $x \in \mathbb{R}^m$ there is a sign vector $s \in \{\pm 1\}^m$ such that $|x| = D_s x$, where $D_s = \mathrm{diag}(s)$.

7.2 A reformulation of CNP and proof of Theorem 2

Lemma 1 ([32]) *Let $A^C \in \mathbb{R}^{n \times p}$, $0 \leq A^\Delta \in \mathbb{R}^{n \times p}$, $w^C \in \mathbb{R}^n$, $0 \leq w^\Delta \in \mathbb{R}^n$. A vector $\xi \in \mathbb{R}^p$ is a solution of a system $A\xi = w$ with some A and w satisfying $A^C - A^\Delta \leq A \leq A^C + A^\Delta$ and $w^C - w^\Delta \leq w \leq w^C + w^\Delta$, if and only if it satisfies $|A^C \xi - w^C| \leq A^\Delta |\xi| + w^\Delta$.* □

Remark. To avoid misunderstanding, we can state the lemma also in the following form: given A^C, A^Δ, w^C, w^Δ as above, it holds

$$\{\xi \in \mathbb{R}^p : \ |A^C \xi - w^C| \leq A^\Delta |\xi| + w^\Delta\} \\ = \bigcup_{\substack{A:\ A^C - A^\Delta \leq A \leq A^C + A^\Delta \\ w:\ w^C - w^\Delta \leq w \leq w^C + w^\Delta}} \{\xi \in \mathbb{R}^p : \ A\xi = w\}.$$

In CNP, we are given (A, w, Γ) and we are to find the minimum number $\delta = \|(\Delta A, \Delta w)\|_{\max}$ such that the linear system $(A + \Delta A)\xi = w + \Delta w$ has a solution ξ and the j-th column of Γ is zero when $j \notin \Gamma$. Let E be a matrix with rows $\eta', \ldots, \eta'$, where η is given by (6) (that is: the j-th column of E is zero when $j \notin \Gamma$, and it is the all-one vector when $j \in \Gamma$). Now CNP can be reformulated as the task to find the minimum δ such that, for a certain A_0 satisfying $A - \delta E \leq A_0 \leq A + \delta E$ and a certain w_0 satisfying $w - \delta e \leq w_0 \leq w + \delta e$, the system $A_0 \xi = w_0$ is solvable. By Lemma 1, this is equivalent to finding the minimum δ such that

$$\mathfrak{B}_\delta := \{\xi : |A\xi - w| \leq \delta E|\xi| + \delta e\} \neq \emptyset. \tag{8}$$

Now we recall (4): $\widehat{\gamma}^n$ and $\widehat{\beta}^n$ are defined as solutions to CNP with data (Z^n, y^n, Γ). We substitute $A := Z^n$ with rows $z'_1, \ldots, z'_n$ and $w := y^n$. We will also write δ^n to emphasize that δ depends on n. We get

$$\begin{aligned}
\mathfrak{B}_{\delta^n} &= \{\xi : |Z^n\xi - y^n| \le \delta^n E|\xi| + \delta^n e\} \\
&= \bigcup_{s\in\{\pm1\}^p} \left\{ \xi : \begin{array}{c} Z^n\xi - y^n \le \delta^n E D_s\xi + \delta^n e, \\ Z^n\xi - y^n \ge -\delta^n E D_s\xi - \delta^n e, \\ D_s\xi \ge 0 \end{array} \right\} && (9) \\
&= \bigcup_{s\in\{\pm1\}^p} \left\{ \xi : \begin{array}{c} \frac{z'_i\xi - y_i}{\eta' D_s\xi + 1} \le \delta^n, \quad i = 1, \ldots, n, \\ \frac{-z'_i\xi + y_i}{\eta' D_s\xi + 1} \le \delta^n, \quad i = 1, \ldots, n, \\ D_s\xi \ge 0 \end{array} \right\}, && (10)
\end{aligned}$$

where we have used the fact that we can write $|\xi| = D_s\xi$ in every orthant $s \in \{\pm1\}^p$ of $\mathbb{R}^p$. Expression (10) shows that finding the minimum $\delta^n \equiv \widehat{\gamma}^n$ such that $\mathfrak{B}_{\delta^n} \neq \emptyset$ is indeed equivalent to (5) and (7). The proof of Theorem 2 is thus complete.

7.3 Geometry

Example 1. In Figure 1 we consider an example motivated by [21] with $n = 3$ and

$$Z^n = \begin{pmatrix} 3 & -0.5 \\ 0.5 & 3 \\ 0.6 & 3 \end{pmatrix}, \qquad y^n = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The Figure depicts (9) where $\delta^n \in \{0.1, 0.2, \ldots, 0.8\}$. It is apparent that $\mathfrak{B}_\delta$ is a union of polyhedra, which are orthant-by-orthant convex. In this example we obviously get $(\delta^n)^* := \min\{\delta^n : \mathfrak{B}_{\delta^n} \neq \emptyset\} = 0$, and the corresponding estimate is $\widehat{\beta}^n = 0$.

Example 2. In Figure 2 we consider the example

$$Z^n = \begin{pmatrix} 3 & -0.5 \\ 0.5 & 3 \\ 0.6 & 3 \end{pmatrix}, \qquad y^n = \begin{pmatrix} 0.2 \\ 0.7 \\ -0.1 \end{pmatrix}.$$

Here, $(\delta^n)^* = 0.323$ and $\widehat{\beta}^n = (0.22, 0.08)'$. The minimum δ^n is attained in the orthant $s = (1, 1)$.

Example 3. Figure 3 depicts what happens with the set $\mathfrak{B}_\delta$ when we add a new observation (z_{n+1}, y_{n+1}). In Figure 3 we can see the set $\mathfrak{B}_{\delta=0.5}$ copied from Figure 1, plus the polyhedron depicting the newly added inequalities in (9):

$$\bigcup_{s\in\{\pm1\}^2} \left\{ \xi : \begin{array}{c} z'_{n+1}\xi - y_{n+1} \le \delta\eta' D_s\xi + \delta, \\ z'_{n+1}\xi - y_{n+1} \ge -\delta\eta' D_s\xi - \delta, \\ D_s\xi \ge 0 \end{array} \right\},$$

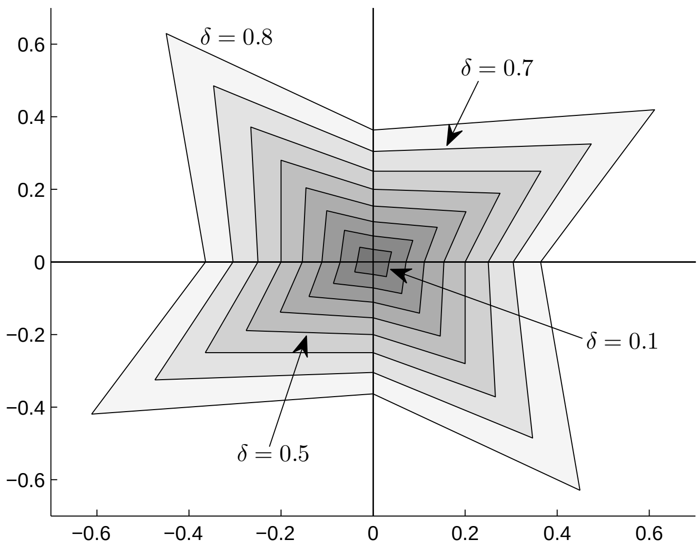

Fig. 1 Example 1 — Sets $\mathfrak{B}_{\delta^n}$ for $\delta^n \in \{0.1, 0.2, \ldots, 0.8\}$.

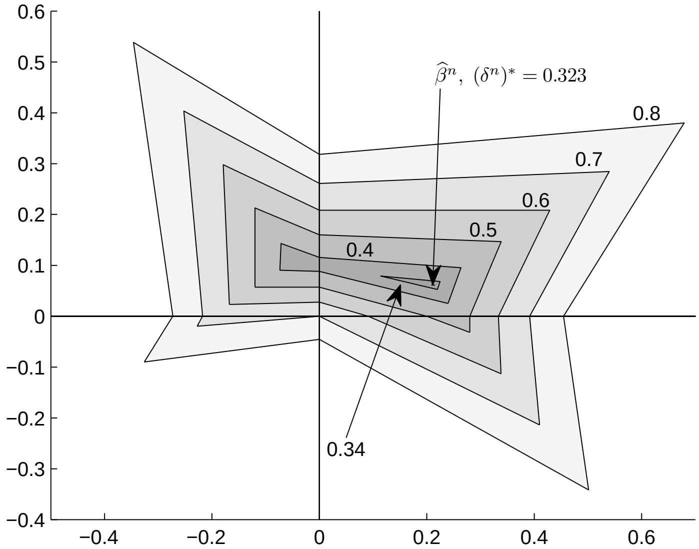

Fig. 2 Example 2 — Sets $\mathfrak{B}_{\delta^n}$ for $\delta^n \in \{0.323, 0.34, 0.4, 0.5,\ 0.6, 0.7, 0.8\}$.

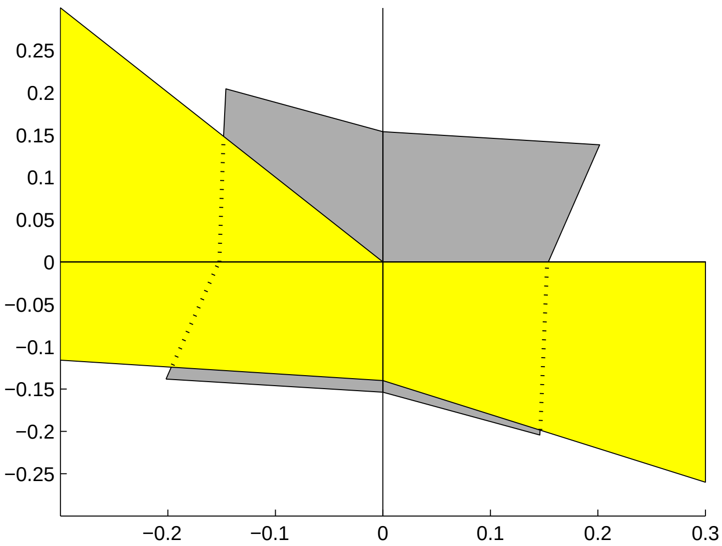

Fig. 3 Example 3 — An additional observation can be understood as a cut of the set $\mathfrak{B}_\delta$.

where $\delta = 0.5$ is fixed, $z'_{n+1} = (0.6, 2.9)$ and $y_{n+1} = 0.05$. Addition of a new observation can thus be understood as a "cut" of $\mathfrak{B}_\delta$.

Adding inequalities may result in infeasibility when δ is kept fixed. If this is the case, it is necessary to increase δ. This is a useful observation which shows:

Lemma 2 *For every n, $\widehat{\gamma}^{n+1} \geq \widehat{\gamma}^n$ a.s.* □

Since we have

$$\widehat{\gamma}^n \leq \gamma \text{ a.s.} \tag{11}$$

(obvious), we conclude that the sequence $\widehat{\gamma}_n$ converges a.s. to a limit γ^*. It remains to show that the undesirable case $\gamma^* < \gamma$ cannot occur. This will be done in the next section. The idea of the proof is based on the fact that Assumption 2 implies that a suitable sequence of cuts exists with a high probability.

7.4 Proof of Theorem 1

We continue with the notation of the previous sections. In particular, from (4) we know that $\widehat{\beta}^n \in \mathfrak{B}_{(\delta^n)^*}$, where $(\delta^n)^*$ is the minimum δ^n such that (8) is nonempty. Now we will study the behavior of $\mathfrak{B}_{(\delta^n)^*}$ when $n \to \infty$. We will show that the probability that the set $\mathfrak{B}_{(\delta^n)^*}$ degenerates to a single point $\{\beta\}$ tends towards 1.

Recall that in Assumption 2 we defined the crucial event $\mathcal{A}_n(\alpha, c, u)$, which was assumed to occur with a high probability when n is large. We will prove that if $\mathcal{A}_n$ holds true, then every point $\widetilde{\beta} \neq \beta$ will be cut off from $\mathfrak{B}_{(\delta^n)^*}$ by a suitable cut, provided that α is sufficiently small. Moreover, convergence of $\widehat{\beta}^n$ to β implies convergence of $\widehat{\gamma}^n$ to γ. The estimators $\widehat{\beta}^n$ and $\widehat{\gamma}^n$ are thus (weakly) consistent.

So, fix an arbitrary $\widetilde{\beta} \neq \beta$ and set

$$u = \widetilde{\beta} - \beta.$$

Choose an α such that

$$0 < \alpha < \frac{c\|u\|}{1 + \eta'|\widetilde{\beta}|} \tag{12}$$

(recall that $|\widetilde{\beta}|$ is the component-wise absolute value of $\widetilde{\beta}$), where $c > 0$ is the number from Assumption 2(i). Also recall also that η was defined in (6). Now assume that $\mathcal{A}_n$ holds true and that $\widetilde{\beta} \in \mathfrak{B}_{(\delta^n)^*}$. Thus $\widetilde{\beta}$ fulfills the system (9). In particular, it fulfills

$$z'_{i_0}\widetilde{\beta} - y_{i_0} \leq (\delta^n)^*\eta' D_s\widetilde{\beta} + (\delta^n)^*, \tag{13}$$

$$z'_{i_0}\widetilde{\beta} - y_{i_0} \geq -(\delta^n)^*\eta' D_s\widetilde{\beta} - (\delta^n)^*, \tag{14}$$

$$D_s\widetilde{\beta} \geq 0$$

for some $s \in \{\pm 1\}^p$ and $i_0 \in \{1, \ldots, n\}$ such that, using Assumption 2,

(i) if $x'_{i_0}u \geq 0$, then:
  (a) $-\varepsilon_{i_0} \geq \gamma - \alpha$,
  (b) $\operatorname{sgn}(\widetilde{\beta}_j)\nu_{i_0 j} \geq \gamma - \alpha$ if $j \in \mathit{\Gamma}$;
(ii) if $x'_{i_0}u < 0$, then:
  (a) $\varepsilon_{i_0} \geq \gamma - \alpha$,
  (b) $-\operatorname{sgn}(\widetilde{\beta}_j)\nu_{i_0 j} \geq \gamma - \alpha$ if $j \in \mathit{\Gamma}$.

We will distinguish between two cases according to the sign of $x'_{i_0}u$. Let $\nu_{i_0} = (\nu_{i_0 1}, \ldots, \nu_{i_0 p})'$.

CASE 1: $x'_{i_0}u \geq c\|u\|$. Using $z_{i_0} = x_{i_0} + \nu_{i_0}$ and $y_{i_0} = x'_{i_0}\beta + \varepsilon_{i_0}$, from (13) we derive

$$(x_{i_0} + \nu_{i_0})'(\beta + u) - ({x_{i_0}}'\beta + \varepsilon_{i_0}) \leq (\delta^n)^*\eta'(\beta + u) + (\delta^n)^*.$$

By rearrangement we get

$$x'_{i_0}u + (\nu'_{i_0} - (\delta^n)^*\eta' D_s)\widetilde{\beta} \leq (\delta^n)^* + \varepsilon_{i_0}. \tag{15}$$

Now

$$
\begin{aligned}
\alpha &\geq (\delta^n)^* + \varepsilon_{i_0} && \text{[using (i)(a) and (11)]}\\
&\geq x'_{i_0}u + (\nu'_{i_0} - (\delta^n)^*\eta' D_s)\widetilde{\beta} && \text{[using (15)]}\\
&\geq c\|u\| + (\nu'_{i_0} - (\delta^n)^*\eta' D_s)\widetilde{\beta} && \text{[CASE 1 assumption]}\\
&\geq c\|u\| + ((\gamma-\alpha)\eta' - (\delta^n)^*\eta')|\widetilde{\beta}| && \text{[using (i)(b)]}\\
&\geq c\|u\| - \alpha\eta'|\widetilde{\beta}|\\
&\geq c\|u\| - \eta'|\widetilde{\beta}|\frac{c\|u\|}{1+\eta'|\widetilde{\beta}|} && \text{[using (12)]}\\
&> \alpha.
\end{aligned}
$$

CASE 2: $x'_{i_0}u \leq -c\|u\|$. From (14) we derive

$$
-(x_{i_0} + \nu_{i_0})'(\beta + u) + (x'_{i_0}\beta + \varepsilon_{i_0}) \leq (\delta^n)^*\eta' D_s\widetilde{\beta} + (\delta^n)^*,
$$

and, after a rearrangement,

$$
-x'_{i_0}u - (\nu_{i_0}{}' + (\delta^n)^*\eta' D_s)\widetilde{\beta} \leq (\delta^n)^* - \varepsilon_{i_0}. \tag{16}
$$

Now, using (ii)(a, b), (11), (12) and (16),

$$
\begin{aligned}
\alpha &\geq (\delta^n)^* - \varepsilon_{i_0}\\
&\geq -x'_{i_0}u - (\nu'_{i_0} + (\delta^n)^*\eta' D_s)\widetilde{\beta}\\
&\geq c\|u\| - (\nu'_{i_0} + (\delta^n)^*\eta' D_s)\widetilde{\beta}\\
&\geq c\|u\| + ((\gamma-\alpha)\eta' - (\delta^n)^*\eta')|\widetilde{\beta}|\\
&\geq c\|u\| - \alpha\eta'|\widetilde{\beta}|\\
&\geq c\|u\| - \eta'|\widetilde{\beta}|\frac{c\|u\|}{1+\eta'|\beta|}\\
&> \alpha.
\end{aligned}
$$

Both CASE 1 and CASE 2 lead to a contradiction. The proof of consistency of $\widehat{\beta}^n$ is now complete. To show that $\widehat{\gamma}^n$ is consistent, assume that $(\delta^n)^* \to \gamma^* < \gamma$. Then, at least one of the inequalities (15) and (16) is violated when $\alpha > 0$ is sufficiently small and n large. This completes the proof of Theorem 1.

8 A generalization: the radii of error distributions need not be the same

So far we have assumed that all error distributions have the same radius γ. It is easy to generalize the theory to the following setup. Let $\gamma_0 > 0$ be an (unknown) radius of the distribution of ε_i and let $\gamma_j \geq 0$ ($j = 1, \ldots, p$) be an (unknown) radius of the error distribution of ν_{ij}. Assume that the *ratios* γ_j/γ_0 are known. Then it suffices to replace the 0-1 vector η defined in (6) by

$$\eta' = \left(\frac{\gamma_1}{\gamma_0}, \ldots, \frac{\gamma_p}{\gamma_0}\right)$$

and the consistency of $\widehat{\beta}^n$ and $\widehat{\gamma}^n \equiv \widehat{\gamma}_0^n$ remains preserved. Thus we also have consistent estimators of γ_j, $j = 1, \ldots, n$. The computational properties are preserved, too.

Remark. This is, in a sense, similar to the classical TLS theory: the knowledge of ratios $\frac{\mathrm{var}(\nu_{ij})}{\mathrm{var}(\varepsilon_i)}$ with $j = 1, \ldots, p$ is known to be, under certain assumptions, a sufficient condition for identification. (Details on identification conditions for EIV models can be found in [6].)

Open problem. The idea of this section leads us to the following interesting question: is it possible to design a consistent estimator of the radii $\gamma_1/\gamma_0, \ldots, \gamma_p/\gamma_0$?

9 Discussion on uniqueness of $\widehat{\boldsymbol{\beta}}^{\boldsymbol{n}}$

Recall that $\widehat{\beta}^n$ is defined as *any* solution to CNP. Although it is unique in the limit $n \to \infty$ in the sense of Theorem 1, for a fixed n it generally need not be unique. Here we exploit the geometry of the problem, showing when it is unique and when it is not. In this section we treat n as a fixed constant.

Recall that $\widehat{\beta}^n$ is defined as *any* point in the polyhedron

$$\mathfrak{P}_s = \{\xi : A_s\xi \leq c_s\},$$

where $s \in \{\pm 1\}^p$ is a suitable sign vector and

$$A_s = \begin{pmatrix} Z^n - (\delta^n)^* E D_s \\ -Z^n - (\delta^n)^* E D_s \\ -D_s \end{pmatrix}, \quad c_s = \begin{pmatrix} y^n + (\delta^n)^* e \\ -y^n + (\delta^n)^* e \\ 0 \end{pmatrix}.$$

Recall also that $\mathfrak{B}_{(\delta^n)^*} = \bigcup_{t \in \{\pm 1\}^p} \mathfrak{P}_t$, see (9).

In Fig. 2 we have an example where $p = 2$, $s = (1, 1)'$ and the polyhedron $\mathfrak{P}_s$ contains a single point $(0.22, 0.08)'$. So in this example $\widehat{\beta}^n$ is unique.

Let us make an easy observation. *If $\widehat{\beta}^n$ is not unique iff*

$$1 \leq \mathrm{affine.dimension}(\mathfrak{P}_s) \leq p - 1. \tag{17}$$

[*Proof.* The second inequality follows from the minimality of $(\delta^n)^*$, see (8), and the first inequality follows from the assumption that $\mathfrak{P}_s$ contains at least two distinct points, and thus also a line segment connecting them (by convexity).]

What can happen is illustrated by example in Fig. 4, where $p = 3$, $s = (1, 1, 1)'$ and affine.dimension$(\mathfrak{P}_s) = 2$. This happens when the system $A_s\xi \leq c_s$ contains two inequalities $\alpha_1'\xi \leq \zeta_1$ and $\alpha_2'\xi \leq \zeta_2$ such that

$$\alpha_1 = -\alpha_2 \quad \text{and} \quad \zeta_1 = -\zeta_2. \tag{18}$$

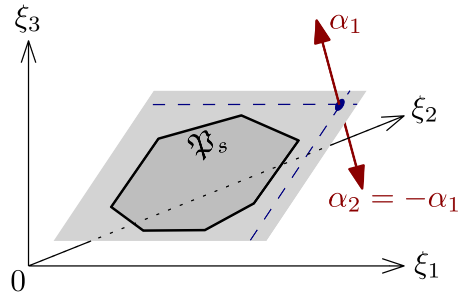

Fig. 4 How $\mathfrak{P}_s$ can look when $\widehat{\beta}^n$ is not unique.

Clearly, $\mathfrak{P}_s$ is a random polyhedron (since the coefficients of the system $A_s\xi \leq c_s$ are random variables). From (17) we can derive the following dimension condition: if the affine dimension of $\mathfrak{P}_s$ is zero a.s., then $\widehat{\beta}^n$ is unique. This holds true, for example, when every p-tuple of distinct inequalities chosen from the system $A_s\xi \leq c_s$ are linearly independent a.s.

The last condition can be expected to be satisfied when the errors are independent and continuously distributed. However, currently we cannot prove it: we cannot rule out such cases as (18). The problem is that the system $A_s\xi \leq c_s$ contains random variables $Z^n, y^n, (\delta^n)^*$, which are dependent, even if we assumed independence of Z^n, y^n. Indeed, $(\delta^n)^*$ depends on both Z^n and y^n.

To conclude, it would be interesting to derive sufficient conditions for a.s. uniqueness of $\widehat{\beta}^n$ when n is fixed. However, this problem is not of high importance, since the possible ambiguity of $\widehat{\beta}^n$ for a particular fixed n does not violate the asymptotic consistency.

10 Conclusions

We have considered an Errors-In-Variables linear regression model with stochastic regressors, where our assumptions are that: (i) it is known which regressors are affected by errors and which are not and; (ii) the matrix of regressors asymptotically fulfills a certain form of regularity; and (iii) all errors share the same bound, which is—roughly stated—approached arbitrarily close with a high probability when the number of observations is sufficiently large. (Then we relaxed the last assumption to the form that error bounds need not be identical, but we need to assume that their *ratios* are known.) We need neither zero means, nor independence of errors, nor identical distributions. We have shown that Total Least Squares, where the Frobenius matrix norm is replaced by the Chebyshev norm, yields a consistent estimator of the parameters. From the computational viewpoint, we have reduced the problem to solving 2^p generalized linear-fractional programming problems, where p is the number of regression parameters. The good news is that computation time of the method is *not* exponential in n, the number of observations, and thus the estimator can be efficiently computed for many practical regression models, where we have at most, say, 20 regression parameters, even if n is large. (2^{20} is still feasible with the aid of today's hardware.) Moreover, the method is easy to parallelize. This is, in a sense, the best possible algorithmic result, since the problem can be shown to be NP-hard.

Acknowledgment. The work was supported by the Czech Science Foundation under grants P402/13-10660S (M. Hladík), P402/12/G097 (M. Černý) and P403/15/09663S (J. Antoch). J. Antoch also acknowledges the support from the BELSPO IAP P7/06 StUDyS network. We are also obliged to Tomáš Cipra, a senior member of DYME Research Center, for fruitful discussions.

References

1. Almeida, H., Campello, M. and Galvao A. F. Jr. (2010) Measurement errors in investment equations. *Review of Financial Studies* **23**, 3279-3328.
2. Bikel, P. J. and Ritov, Y. (1987) Efficient estimation of in the errors in variables model. *Annals of Statistics* **15**, 513-540.
3. Biorn, E. (2010) Panel data with measurement errors: Instrumental variables and GMM procedures combining levels and differences. *Econometric Reviews* **19**, 391-424.
4. Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. (1997) *Measurement Error in Nonlinear Models: A Modern Perspective, 2nd ed.* CRC, Boca Raton.
5. Chen, H.-Y., Lee, A. and Lee, C.-F. (2015) Alternative errors-in-variables models and their applications in finance research. *The Quarterly Review of Economics and Finance* **58**, 213-227.
6. Cheng, C. I. and Van Ness, J. W. (1999) *Regression with Measurement Error.* Oxford University Press, New York.
7. Dagenais, M. and Dagenais, D. (1997) Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics* **76**, 193-222.
8. Durbin, J. (1954) Errors in variables. *Review of the International Statistical Institute* **22**, 23-32.
9. Erickson, T. (2001) Constructing instruments for regressions with measurement error when no additional data are available: Comment. *Econometrica* **69**, 221-222.
10. Erickson, T., Jiang, C. H. and Whited, T. (2014) Minimum distance estimation of the errors-in-variables model using linear cumulant equations. *Journal of Econometrics* **183**, 211-221.
11. Erickson, T. and Whited, T. (2000) Measurement error and the relationship between investment and q. *Journal of Political Economy* **108**, 1027-1057.
12. Erickson, T. and Whited, T. (2002) Two-step GMM estimation of the errors-in-variables model using higher-order moments. *Econometric Theory* **18**, 776-799.
13. Freund, R. W. and Jarre, F. (1995) An interior-point method for multifractional programs with convex constraints. *Journal of Optimization Theory and Applications* **85** (1), 125–161.
14. Gallo, P. P. (1982a) *Properties of Estimators in Errors-in-Variables Models.* PhD thesis, University of North Carolina, Chapel Hill.
15. Gallo, P. P. (1982b) Consistency of regression estimates when some variables are subject to error. *Communications in Statistics – Theory and Methods* **11**, 973-983.
16. Gleser, L. J. (1981) Estimation in multivariate "errors in variables" regression models. Large sample results. *Annals of Statistics* **9**, 22-44.
17. Golub, G. H. and Van Loan, C. F. (1980) An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis* **17**, 883-893.
18. Golub, G. H., Hoffman, A. and Stewart, G. W. (1987) A generalization of the Eckart-Young-Mirsky matrix approximation theorem. *Linear Algebra and Its Applications* **88**, 322-327.

19. Griliches, Z. and Hausman, J. A. (1986) Errors in variables in panel data. *Journal of Econometrics* **31**, 93-118.
20. Healy, J. D. (1975) *Estimation and tests for unknown linear restrictions in multivariate linear models.* PhD thesis, Purdue University.
21. Hladík, M. and Černý M. (2015) Total Least Squares and Chebyshev norm. *Procedia Computer Science* **51**, 1791-1800.
22. Küchenhoff, H. (1995) The identification of logistic regression models with errors in the variables. *Statistical Papers* **36** (1), 41-47.
23. Kumunjer, I. and Ng, S. (2014) Measurement errors in dynamic models. *Econometric Theory* **30**, 150-175.
24. Kukush, A., Markovsky, I. and Van Huffel, S. (2002) On consistent estimators in linear and bilinear multivariate Errors-In-Variables models. In Van Huffel, S. and Lemmerling, P., eds. *Total Least Squares and Errors-In-Variables Modeling: Analysis, Algorithms and Applications*, SIAM, Philadelphia, 155-164.
25. Kukush, A. and Van Huffel, S. (2004) Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model $AX = B$. *Metrika* **59**, 75-97.
26. Kukush, A., Markovsky, I. and Van Huffel, S. (2005) Consistency of the structured total least squares estimator in a multivariate errors-in-variables model. *Journal of Statistical Planning and Inference* **133(2)**, 315-358.
27. Lewbel, A. (1997) Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R& D. *Econometrica* **65**, 1201-1213.
28. Maddala, G. S. and Nimalendran, M. (1996) Error-in-variables problems in financial models. In: Maddala, G. S. and Rao, C.R. (Eds.), *Handbook of Statistics*, Elsevier Science, New York.
29. Markovsky, I., Rastellob, M. L., Premolic, A., Kukush, A. and Van Huffel, S. (2005) The element-wise weighted total least-squares problem. *Computational Statistics & Data Analysis* **50**, 181-209.
30. Nesterov, Y. E. and Nemirovski, A. S. (1995) An interior-point method for generalized linear-fractional programming. *Mathematical Programming* **69** (1B), 177–204.
31. Okamoto, M. (1973) Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Annals of Statistics* **1**, 763-765.
32. Oettli, W. and Prager, W. (1964) Compatibility of approximate solution of linear equations with given error bounds for coefficients and right-hand sides. *Numerische Mathematik*, **6**, 405–409.
33. Pešta, M. (2011) Strongly consistent estimation in dependent errors-in-variables. *AUC – Mathematica et Physica* **52**, 69-79.
34. Pešta, M. (2016) Unitarily invariant errors-in-variables estimation. *Statistical Papers.* First Online, DOI:10.1007/s00362-016-0800-9.
35. Rao, C. R. (1980) Matrix approximations and reduction of dimensionality in multivariate statistical analysis. *Multivariate Analysis **5***, Krishnaiah, P.R. ed., Amsterdam. North-Holland, 3-22.
36. Shalabh, Garg, G. and Misra, N. (2010) Consistent estimation of regression coefficients in ultrastructural measurement error model using stochastic prior information. *Statistical Papers* **51** (3), 717-748.
37. Singh S., Jain K. and Sharma S. (2014) Replicated measurement error model under exact linear restrictions. *Statistical Papers* **55** (2), 253-274.
38. Tang, N. S., Li, D. W. and Tang, A. M. (2016) Semiparametric Bayesian inference on generalized linear measurement error models. *Statistical Papers.* First Online, DOI:10.1007/s00362-016-0739-x.
39. Van Huffel, S. and Vandewalle, J. (1991) *The Total Least Squares Problem: Computational Aspects and Analysis.* SIAM, Philadelphia.
40. von Neumann, J. (1937) Some matrix inequalities and metrization of matric-space. *Tomsk University Review* **1**, 286-300.