Noname manuscript No.
(will be inserted by the editor)

Distributionally robust L_1-estimation in multiple linear regression

Zhaohua Gong · Chongyang Liu · Jie Sun · Kok Lay Teo

Received: date / Accepted: date

Abstract Linear regression is one of the most important and widely used techniques in data analysis, for which a key step is the estimation of the unknown parameters. However, it is often carried out under the assumption that the full information of the error distribution is available. This is clearly unrealistic in practice. In this paper, we propose a distributionally robust formulation of L_1-estimation (or the least absolute value estimation) problem, where the only knowledge on the error distribution is that it belongs to a well-defined ambiguity set. We then reformulate the estimation problem as a computationally tractable conic optimization problem by using duality theory. Finally, a numerical example is solved as a conic optimization problem to demonstrate the effectiveness of the proposed approach.

Keywords Multiple linear regression · Least absolute value estimation · Conic optimization · Semi-infinite optimization

1 Introduction

Linear regression is one of the most important and widely used techniques in data analysis [1], for which a key step is the estimation of the unknown parameters. Traditionally, it is formulated based on the principle of least squares, where the model parameters are to be chosen such that the sum of squares of the distances between the observations and the fitting line is minimized subject to the assumptions that the errors are normally distributed and are homoscedastic. Under these assumptions, the linear least squares estimation produces the best estimates in terms of linear unbiased estimation and maximum likelihood estimation [2]. However, the results obtained using linear least squares regression tend to be very sensitive to outliers. To address this problem, robust regression methods are proposed by some researchers, where different functions of residuals, instead of least squares function, are introduced for

Z. Gong
School of Mathematics and Information Science, Shandong Institute of Business and Technology, Yantai, China
E-mail: zhaohuagong@163.com
C. Liu (✉)
School of Mathematics and Information Science, Shandong Institute of Business and Technology, Yantai, China
E-mail: liu_chongyang@yahoo.com
J. Sun
School of Electrical Engineering, Computing, and Mathematical Sciences, Curtin University, Perth, Australia
E-mail: jie.sun@curtin.edu.au
K.L. Teo
School of Electrical Engineering, Computing, and Mathematical Sciences, Curtin University, Perth, Australia
E-mail: k.l.teo@curtin.edu.au

minimization. The least absolute value (LAV) method is one of the most popular robust regression methods and it results in the maximum-likelihood estimation of regression parameters given a double exponentially distributed error [3–5]— an alternative to the normal error distribution with a wide range of applications, for example, in autoregression model [6] and stock market returns [7]. Unlike the least squares regression, it has been shown that LAV regression does not have a close-form solution, and the solution may not be unique [8]. By reformulating the LAV regression as a linear programming problem, many efficient algorithms are available in the literature to solve the problem; see, for example [9–12] and a comprehensive survey [13]. Although the methods mentioned above are interesting, they all assume that the full information on the distribution of the error in their regression model is available, which is often unrealistic in practice.

In this paper, we consider a multiple linear regression model without exact knowledge on the distribution of the error. The knowledge that we have on the distribution of the error is that it belongs to an ambiguity set defined by certain statistic information. Furthermore, the error in the regression model is not necessarily homoscedastic. Under these circumstances, we propose a distributionally robust least absolute value estimation (DR-LAVE) formulation. Distributionally robust optimization has gained significant interest in recent years. For example, distributionally robust quantile optimization problem is discussed in [14]. General conditions for polynomial time solvability of a generic distributionally robust model are given in [15]. Tractable approximations to two-stage and multistage distributionally linear programs are derived in [16–19]. A model and an algorithm for distributionally robust least squares problem are studied in [20]. In this paper, we first show that the inner optimization problem, which involves the worst-case expectation value, is equivalent to a semi-infinite programming problem. Then, the semi-infinite constraints in the equivalent problem are further reformulated as a conic optimization problem using duality theory. On this basis, the DR-LAVE problem is reformulated as a computationally tractable conic optimization problem. Finally, a numerical example is presented to demonstrate the effectiveness of the proposed method.

Notations. We denote a random vector, say $\tilde{z}$, with the tilde sign. Matrices and vectors are represented as upper and lower case letters, respectively. In particular, e is a vector of all ones in $\mathbb{R}^n$. The set $\mathcal{P}_0(\mathbb{R}^n)$ represents the space of probability distributions on $\mathbb{R}^n$. If $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n \times \mathbb{R}^t)$ is a joint probability distribution of two random vectors $\tilde{z} \in \mathbb{R}^n$ and $\tilde{u} \in \mathbb{R}^t$, then $\Pi_{\tilde{z}}\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n)$ denotes the marginal distribution of $\tilde{z}$ under $\mathbb{P}$. This definition is extended to ambiguity set $\mathcal{P} \subseteq \mathcal{P}_0(\mathbb{R}^n \times \mathbb{R}^t)$ by setting $\Pi_{\tilde{z}}\mathcal{P} = \cup_{\mathbb{P}\in\mathcal{P}}\{\Pi_{\tilde{z}}\mathbb{P}\}$. For a proper cone $\mathcal{K}$ (i.e., a closed, convex, and pointed cone with nonempty interior) in a finite dimensional Hilbert space, the relation $x \preceq_{\mathcal{K}} y$ indicates that $y - x \in \mathcal{K}$. Finally, the dual cone of $\mathcal{K}$ is denoted by

$$\mathcal{K}^* := \{y : \langle y, x\rangle \geq 0, \forall x \in \mathcal{K}\},$$

where $\langle \cdot, \cdot \rangle$ is the inner product.

2 Problem formulation

Consider the following multiple linear regression model:

$$z_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_m x_{i,m} + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where z_i is the ith observation on the dependent variable; $x_{i,j}$, $j = 1, 2, \ldots, m$, are the ith observations on the independent variables; β_j, $j = 0, 1, \ldots, m$, are the regression coefficients to be estimated; and ε_i is the ith error. Let $z = (z_1, z_2, \ldots, z_n)^\top$, $\beta = (\beta_0, \beta_1, \ldots, \beta_m)^\top$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top$, and

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{pmatrix}. \tag{2}$$

Then, model (1) can be written in matrix form as given below:

$$z = X\beta + \varepsilon. \tag{3}$$

Let $\tilde{z}$ be the random sample observation vector. Then, the least absolute value residual can be defined as

$$\|X\beta - \tilde{z}\|_1 := \sum_{k=1}^{n} |x^k \beta - \tilde{z}_k|, \tag{4}$$

where x^k is the kth row of the matrix X. In general, the information on the distribution of the error vector ε is not known exactly. That is, we do not know the exact distribution $\mathbb{P}$ of $\tilde{z}$. However, it is assumed that $\mathbb{P}$ belongs to an ambiguity set $\mathcal{P}$ of distributions. Accordingly, instead of calculating $\mathbb{E}_{\mathbb{P}}\{\|X\beta - \tilde{z}\|_1\}$, we calculate $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\{\|X\beta - \tilde{z}\|_1\}$, where $\mathbb{E}$ stands for the expectation; in other words, we use the worst-case expected L_1-norm objective function over $\mathcal{P}$. Thus, our DR-LAVE problem can be formulated as

$$\text{(DR-LAVE)} \qquad \min_{\beta} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\{\|X\beta - \tilde{z}\|_1\}. \tag{5}$$

Problem DR-LAVE is, in essence, a min-max stochastic optimization problem, where the inner optimization is a maximization of expectation over a probability measure set of infinite dimension. However, it has been shown that the calculation of the expectation in the inner optimization poses numerical challenge [21]. In the next section, by using the specification of an ambiguity set, we will reformulate the inner optimization problem as a conic optimization problem such that the DR-LAVE becomes computationally tractable.

3 A tractable reformulation of (DR-LAVE)

For the DR-LAVE problem, the ambiguity set $\mathcal{P}$ needs to be well-defined. There exist different specifications of the ambiguity set; see, for example [15, 16, 22, 23]. In particular, a very general format for the ambiguity set is introduced in [23]. This format uses expectation constraint as a basic building block. Motivated by [23], we assume that the ambiguity set $\mathcal{P}$ in our DR-LAVE is represented in the form of

$$\mathcal{P} := \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n \times \mathbb{R}^t) : \mathbb{E}_{\mathbb{P}}\{E\tilde{z} + F\tilde{u}\} = g, \mathbb{P}((\tilde{z}, \tilde{u}) \in \Omega) = 1\}, \tag{6}$$

where $\mathbb{P}$ represents a joint probability distribution of the random vector $\tilde{z} \in \mathbb{R}^n$ in DR-LAVE and some auxiliary random vector $\tilde{u} \in \mathbb{R}^t$, $E \in \mathbb{R}^{p\times n}$, $F \in \mathbb{R}^{p\times t}$ and $g \in \mathbb{R}^p$. Furthermore, the set Ω is of full dimension, compact and representable by a conic inequality

$$\Omega := \{(z, u) : Gz + Hu \preceq_{\mathcal{K}} h\}, \tag{7}$$

with $G \in \mathbb{R}^{r\times n}$, $H \in \mathbb{R}^{r\times t}$, $h \in \mathbb{R}^r$ and $\mathcal{K}$ being a proper cone. It is noted that we allow F and H to be zero matrices, in which case the auxiliary vector $\tilde{u}$ is absent.

Note that the above ambiguity set is less general than the ambiguity set defined in [23]. However, it is general enough to cover many applications of DR-LAVE, e.g., the case of Ω being a box or an ellipse. Under this form of the set Ω, we have the following theorem.

Theorem 1 *Assume that the inner optimization problem*

$$\begin{aligned} \sup_{\mathbb{P}} \ & \mathbb{E}_{\mathbb{P}}\{\|X\beta - \tilde{z}\|_1\} \\ \text{s.t.}\ & \mathbb{E}_{\mathbb{P}}\{E\tilde{z} + F\tilde{u}\} = g \\ & \mathbb{P}((\tilde{z}, \tilde{u}) \in \Omega) = 1 \end{aligned} \tag{8}$$

is bounded. Then, problem (8) is equivalent to the following semi-infinite optimization problem

$$\begin{aligned}
\min_{\gamma,\eta,w}\ & g^\top\gamma+\eta \\
\text{s.t.}\ & (Ez+Fu)^\top\gamma+\eta \geq e^\top w,\ \forall (z,u)\in\Omega \\
& w \geq (X\beta - z),\ \forall (z,u)\in\Omega \\
& w \geq -(X\beta - z),\ \forall (z,u)\in\Omega \\
& \gamma\in\mathbb{R}^p, \eta\in\mathbb{R}, w\in\mathbb{R}^n.
\end{aligned} \tag{9}$$

Proof The inner optimization problem is an optimization problem with respect to a probability measure $\mathbb{P}\in\mathcal{P}_0(\mathbb{R}^n\times\mathbb{R}^t)$ on support Ω. For convenience of discussion, we only consider absolutely continuous random vectors. With this, the expectation in (8) can be written as Lebesgue integrals. Furthermore, $\|X\beta-\tilde{z}\|_1$ is a convex, continuous, and proper (i.e., nowhere$=+\infty$ and not everywhere $=-\infty$) function in β.

Now, by recalling the definitions of $\mathcal{P}$ and Ω in (6) and (7), the inner optimization problem of DR-LAVE can be written as the following problem:

$$\begin{aligned}
\max\ & \int_\Omega \|X\beta - z\|_1 d\mathbb{P}(z,u) \\
\text{s.t.}\ & \int_\Omega (Ez+Fu)d\mathbb{P}(z,u) = g \\
& \int_\Omega \mathbf{1}_{[(z,u)\in\Omega]} d\mathbb{P}(z,u) = 1 \\
& \mathbb{P}\in\mathcal{P}_0(\mathbb{R}^n\times\mathbb{R}^t).
\end{aligned} \tag{10}$$

The dual of the problem (10) is given by

$$\begin{aligned}
\min_{\gamma,\eta}\ & g^\top\gamma+\eta \\
\text{s.t.}\ & (Ez+Fu)^\top\gamma+\eta \geq \|X\beta - z\|_1,\ \forall (z,u)\in\Omega \\
& \gamma\in\mathbb{R}^p, \eta\in\mathbb{R}.
\end{aligned} \tag{11}$$

To argue that the strong duality between (10) and (11) is valid, we note that, as a continuous function of (z,u) over the compact set Ω, the function $\|X\beta-z\|_1$ is bounded above over Ω, namely, there exists a constant $C\geq 0$, such that

$$\|X\beta - z\|_1 \leq C, \quad \forall (z,u)\in\Omega. \tag{12}$$

Thus, $(\gamma,\eta)=(0,C)$ is a strict feasible solution of (11), and the optimal value of (11) is bounded. This shows that the strong duality between (10) and (11) holds according to Theorem 17 in [24].

Next, for each β, let

$$\mathcal{F}_1(\beta) := \{(\gamma,\eta)\in\mathbb{R}^p\times\mathbb{R} : (Ez+Fu)^\top\gamma+\eta \geq \|X\beta-z\|_1\}$$

$$\mathcal{F}_2(\beta) := \{(\gamma,\eta,w)\in\mathbb{R}^p\times\mathbb{R}\times\mathbb{R}^n : (Ez+Fu)^\top\gamma+\eta\geq e^\top w,\ w\geq \pm(X\beta - z),\ \forall (z,u)\in\Omega\}$$

and its projection on the $\gamma-\eta$ space

$$\bar{\mathcal{F}}_2(\beta) := \{(\gamma,\eta)\in\mathbb{R}^p\times\mathbb{R} : \exists w\in\mathbb{R}^n \text{ such that } (\gamma,\eta,w)\in\mathcal{F}_2(\beta)\}.$$

Then, $\mathcal{F}_1(\beta)=\bar{\mathcal{F}}_2(\beta)$, $\forall\beta$. As a result, problem (11) is equivalent to problem (9) since the objective function $g^\top\gamma+\eta$ is independent of w. The proof is complete.

From Theorem 1, we see that the inner optimization problem of DR-LAVE can be solved by solving a semi-infinite optimization problem. However, the first three constraints in problem (9) are semi-infinite constraints, which are difficult to deal with numerically [25]. In the sequel, we will show that these semi-infinite constraints can be transformed into some conic constraints.

Theorem 2 *The semi-infinite constraint*

$$(Ez + Fu)^\top \gamma + \eta \geq e^\top w, \ \forall (z, u) \in \Omega \tag{13}$$

is satisfied if and only if there is a $\phi \in \mathcal{K}^*$ *such that* $h^\top \phi + \eta - e^\top w \geq 0$, $G^\top \phi = E^\top \gamma$ *and* $H^\top \phi = F^\top \gamma$.

Proof The semi-infinite constraint (13) is equivalent to

$$(Ez + Fu)^\top \gamma + \eta - e^\top w \geq 0, \ \forall (z, u) \in \mathbb{R}^n \times \mathbb{R}^t : Gz + Hu \preceq_{\mathcal{K}} h.$$

This constraint is satisfied if and only if the optimal value of the following problem

$$\begin{aligned} \min_{z,u} \ & (Ez + Fu)^\top \gamma + \eta - e^\top w \\ \text{s.t.} \ & Gz + Hu \preceq_{\mathcal{K}} h \\ & z \in \mathbb{R}^n, u \in \mathbb{R}^t \end{aligned} \tag{14}$$

is greater than zero. The dual problem of (14) is given by

$$\begin{aligned} \max_{\phi} \ & h^\top \phi + \eta - e^\top w \\ \text{s.t.} \ & G^\top \phi = E^\top \gamma \\ & H^\top \phi = F^\top \gamma \\ & \phi \in \mathcal{K}^*. \end{aligned} \tag{15}$$

Since the support set Ω defined in (7) is of full dimension, problem (14) satisfies the Slater's condition. Furthermore, since Ω is compact, the optimal value of (14) is finite. Thus, the strong duality of conic optimization (see, Theorem A.2.1 [26]) holds, and (13) is valid if and only if the optimal value of problem (15) is greater than zero, which is, in turn, equivalent to the following system

$$\begin{cases} h^\top \phi + \eta - e^\top w \geq 0 \\ G^\top \phi = E^\top \gamma \\ H^\top \phi = F^\top \gamma \\ \phi \in \mathcal{K}^* \end{cases}$$

being feasible. This proves the theorem.

We next show that the second constraint in problem (9) is equivalent to a conic constraint in the following theorem.

Theorem 3 *The semi-infinite constraint*

$$w \geq (X\beta - z), \ \forall (z, u) \in \Omega \tag{16}$$

is satisfied if and only if there are $\psi_j \in \mathcal{K}^*$, $j = 1, 2, \ldots, n$, *such that* $h^\top \psi_j + w_j - x^j \beta \geq 0$, $G^\top \psi_j = e_j$ *and* $H^\top \psi_j = 0$.

Proof The semi-infinite constraint (16) is equivalent to

$$w_j - x^j \beta + z_j \geq 0,\ \forall j \in \{1, 2, \dots, n\},\ \forall (z, u) \in \mathbb{R}^n \times \mathbb{R}^t : Gz + Hu \preceq_{\mathcal{K}} h.$$

This constraint is satisfied if and only if the respective optimal values of the following n problems

$$\begin{aligned} \min_{z,u}\ & w_j - x^j \beta + z_j \\ \text{s.t.}\ & Gz + Hu \preceq_{\mathcal{K}} h \\ & z \in \mathbb{R}^n, u \in \mathbb{R}^t \end{aligned} \tag{17}$$

are greater than zero. Since the support set Ω defined in (7) is of full dimension, problem (17) satisfies the Slater's condition. Furthermore, since Ω is compact, the optimal value of (17) is finite. Thus, the strong duality of conic optimization holds. As a result, the optimal values of problem (17) are greater than zero if and only if the respective optimal values of the following n dual problems

$$\begin{aligned} \max_{\psi_j}\ & h^\top \psi_j + w_j - x^j \beta \\ \text{s.t.}\ & G^\top \psi_j = e_j \\ & H^\top \psi_j = 0 \\ & \psi_j \in \mathcal{K}^* \end{aligned} \tag{18}$$

are greater than zero. This implies that the following system

$$\begin{cases} h^\top \psi_j + w_j - x^j \beta \geq 0,\ j = 1, 2, \dots, n \\ G^\top \psi_j = e_j,\ j = 1, 2, \dots, n \\ H^\top \psi_j = 0,\ j = 1, 2, \dots, n \\ \psi_j \in \mathcal{K}^*,\ j = 1, 2, \dots, n \end{cases}$$

is feasible, as required.

Similar to Theorem 3, we have the following result.

Corollary 1 *The semi-infinite constraint*

$$w \geq -(X\beta - z),\ \forall (z, u) \in \Omega \tag{19}$$

is satisfied if and only if there are $\varphi_j \in \mathcal{K}^*$, $j = 1, 2, \dots, n$, *such that* $h^\top \varphi_j + w_j + x^j \beta \geq 0$, $G^\top \varphi_j = -e_j$ *and* $H^\top \varphi_j = 0$.

Now, we are ready to provide the main result showing that DR-LAVE can be reformulated as a computationally tractable conic optimization problem.

Theorem 4 *The distributionally robust L_1-estimation problem with the ambiguity set (6) can be reformulated as the following conic optimization problem*

$$
\begin{aligned}
\min_{\beta,\gamma,\eta,w,\phi,\psi,\varphi} \quad & g^\top \gamma + \eta \qquad (20)\\
\text{s.t.} \quad & H^\top \phi = F^\top \gamma \\
& G^\top \phi = E^\top \gamma \\
& h^\top \phi + \eta - e^\top w \geq 0 \\
& h^\top \psi_j + w_j - x^j \beta \geq 0, \ j = 1,2,\ldots,n \\
& h^\top \varphi_j + w_j + x^j \beta \geq 0, \ j = 1,2,\ldots,n \\
& G^\top \psi_j = e_j, \ j = 1,2,\ldots,n \\
& H^\top \psi_j = 0, \ j = 1,2,\ldots,n \\
& G^\top \varphi_j = -e_j, \ j = 1,2,\ldots,n \\
& H^\top \varphi_j = 0, \ j = 1,2,\ldots,n \\
& \gamma \in \mathbb{R}^p, \eta \in \mathbb{R}, w \in \mathbb{R}^n, \phi, \psi_j, \varphi_j \in \mathcal{K}^*, \ j = 1,2,\ldots,n.
\end{aligned}
$$

Proof Combining Theorems 1-3 together with Corollary 1 yields (20).

From Theorem 4, we see that DR-LAVE problem with the ambiguity set (6) can be reformulated as a computationally tractable conic optimization problem (20). Note that ambiguity sets of type (6) offer striking modeling power in spite of the simplicity of expectation and support conditions. In particular, our well-defined ambiguity set allows us to encode (full or partial) information about certain higher-order moments of $\tilde{z}$. For example, assume that the ambiguity set is

$$\mathcal{Q} = \{\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^n) : \mathbb{E}_\mathbb{Q}\{\tilde{z}\} = \mu, \ \mathbb{E}_\mathbb{Q}\big\{(\tilde{z}-\mu)(\tilde{z}-\mu)^\top\big\} \preceq \varSigma, \mu \in \mathbb{R}^n, \varSigma \in \mathbb{S}^n_+\}, \qquad (21)$$

where $\mathbb{S}^n_+$ is the cone of positive semidefinite matrices in $\mathbb{R}^{n\times n}$. We first introduce an auxiliary random matrix $\tilde{U} \in \mathbb{R}^{n\times n}$ such that

$$\mathbb{E}_{\mathbb{Q}'}\{\tilde{U}\} = \varSigma$$

and

$$(\tilde{z}-\mu)(\tilde{z}-\mu)^\top \preceq \tilde{U} \qquad (22)$$

hold almost surely. Then, by Schur's complement lemma [27], equation (22) holds if and only if

$$\begin{bmatrix} 1 & (\tilde{z}-\mu)^\top \\ (\tilde{z}-\mu) & \tilde{U} \end{bmatrix} \succeq 0$$

holds almost surely, which implies

$$\mathbb{P}\left(\begin{bmatrix} 1 & (\tilde{z}-\mu)^\top \\ (\tilde{z}-\mu) & \tilde{U} \end{bmatrix} \succeq 0 \right) = 1, \qquad (23)$$

where $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n \times \mathbb{R}^{n\times n})$. Thus, by applying lifting theorem (see, Theorem 5 [23]) to

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n \times \mathbb{R}^{n\times n}) : \begin{array}{c} \mathbb{E}_\mathbb{P}\{\tilde{z}\} = \mu, \ \mathbb{E}_\mathbb{P}\{\tilde{U}\} = \varSigma, \\ \mathbb{P}\left(\begin{bmatrix} 1 & (\tilde{z}-\mu)^\top \\ (\tilde{z}-\mu) & \tilde{U} \end{bmatrix} \succeq 0 \right) = 1 \end{array} \right\}, \qquad (24)$$

we obtain that $\mathcal{Q} = \varPi_{\tilde{z}}\mathcal{P}$, and $\mathcal{P}$ is an instance of the standardized ambiguity set (6). Therefore the ambiguity set (21) can be replaced by set (24) in a DR-LAVE model.

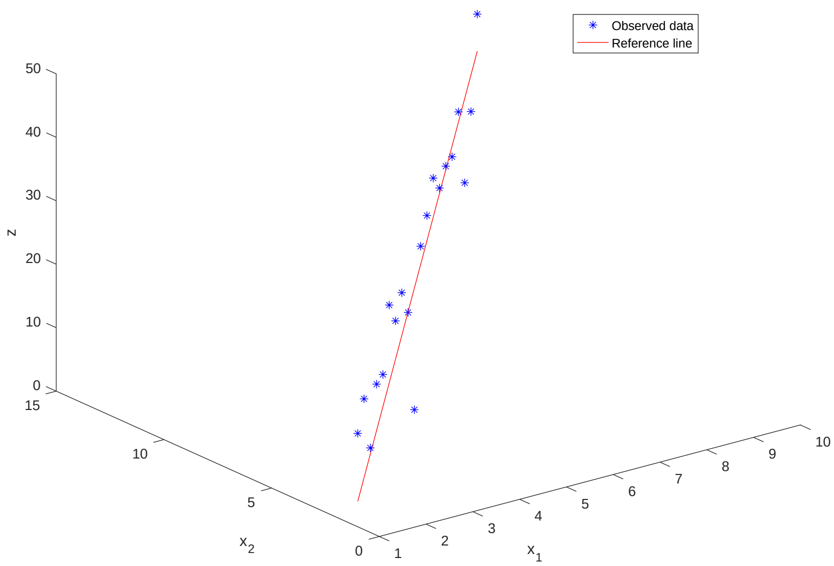

Fig. 1 The observed data and reference line for numerical example.

4 Numerical example

Consider a multiple linear regression model:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon. \tag{25}$$

Assume that we have 20 observations of the independent and dependent variables. In particular, the independent variables are taken as

$$x_1 = \frac{9}{19}i + \frac{10}{19}, \quad x_2 = \frac{14}{19}i + \frac{5}{19}, \quad i = 1, 2, \ldots, 20. \tag{26}$$

To generate the observations of dependent variables, we take $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$ as the nominal values of regression coefficients and assume that the error $\varepsilon_i = \bar{\varepsilon}_i - \mathbb{E}\{\bar{\varepsilon}_i\}$, where $\bar{\varepsilon}_i$ follows a beta distribution with parameters 2 and 5. The generated observations are shown in Fig. 1.

To solve the corresponding conic optimization problem, the required moment information (obtained by 100,000 random samplings) and bounds on the observations of dependent variables are listed in Table 1. Namely, the vector g in ambiguity set (6) is taken as the expectation values in Table 1, and

$$h = \begin{bmatrix} -l_1 \\ l_2 \end{bmatrix}.$$

Moreover, $\mathcal{K} \in \mathbb{R}^{40}_{+}$, and the matrices

$$E = I_{20},\ F = 0_{20},\ G = \begin{bmatrix} -I_{20} \\ I_{20} \end{bmatrix},\ H = \begin{bmatrix} 0_{20} \\ 0_{20} \end{bmatrix},$$

Table 1 Moment information and bounds on the observations of dependent variables for numerical example.

Observation vector ($\tilde{z}$)	$\tilde{z}_1$	$\tilde{z}_2$	$\tilde{z}_3$	$\tilde{z}_4$	$\tilde{z}_5$
Expectation value ($\mathbb{E}\{\tilde{z}\}$)	3.96776	5.66768	7.36542	9.05039	10.68321
Lower bound (l_1)	−2.80583	−2.65548	−2.36696	−2.25372	−2.18161
Upper bound (l_2)	182.67284	199.18573	216.28375	233.07215	249.86035
Observation vector ($\tilde{z}$)	$\tilde{z}_6$	$\tilde{z}_7$	$\tilde{z}_8$	$\tilde{z}_9$	$\tilde{z}_{10}$
Expectation value ($\mathbb{E}\{\tilde{z}\}$)	12.41082	14.13483	15.78565	17.48911	19.14914
Lower bound (l_1)	−2.09822	−1.82269	−1.68129	−1.63922	−1.32288
Upper bound (l_2)	266.79859	283.37665	300.40411	317.40813	333.57384
Observation vector ($\tilde{z}$)	$\tilde{z}_{11}$	$\tilde{z}_{12}$	$\tilde{z}_{13}$	$\tilde{z}_{14}$	$\tilde{z}_{15}$
Expectation value ($\mathbb{E}\{\tilde{z}\}$)	20.85654	22.53802	24.23189	25.89142	27.62000
Lower bound (l_1)	−1.08783	−0.88735	−0.81495	−0.54345	−0.53931
Upper bound (l_2)	351.20721	367.97705	384.32908	401.49039	418.52833
Observation vector ($\tilde{z}$)	$\tilde{z}_{16}$	$\tilde{z}_{17}$	$\tilde{z}_{18}$	$\tilde{z}_{19}$	$\tilde{z}_{20}$
Expectation value ($\mathbb{E}\{\tilde{z}\}$)	29.27940	30.95616	32.63076	34.31396	35.96615
Lower bound (l_1)	−0.31453	−0.14106	−0.03699	0.20438	0.35106
Upper bound (l_2)	435.18753	452.26548	468.44136	485.62086	502.04412

where I_{20} is a 20×20 identity matrix; and 0_{20} is a 20×20 zero matrix. We use the standard software package CVX described in [28] with subroutine SeDuMi to solve the resulting conic optimization problem. The obtained optimal estimates are $\beta_0 = 4.22215 \times 10^{-3}$, $\beta_1 = 1.17934 \times 10^{-2}$, and $\beta_2 = 2.33526$, and it takes 0.67 seconds on Matlab 2016b platform in a personal computer with Intel Core i7-6700 CPU 3.40GHz and 8GB of memory. The corresponding regression line is plotted in Fig. 2. For comparison, we also respectively solve the least squares regression and the LAV regression with the observations in Fig. 1. The obtained least squares estimates are $\beta_0 = 11.02094$, $\beta_1 = 0$, and $\beta_2 = 1.39331$, and the obtained LAV estimates are $\beta_0 = 11.99012$, $\beta_1 = 3.51145$, and $\beta_2 = -0.98259$. The corresponding regression lines are also shown in Fig. 2. From Fig. 2, we can see that our distributionally robust LAV regression line is closer to the reference line than the lines of the least squares and normal LAV regressions. More important, our distributionally robust LAV regression does not require the exact distribution information of the observed data.

5 Conclusion

In this paper, we considered least absolute value estimation for multiple linear regression with non-exact error distribution. We proposed a distributionally robust formulation of the estimation problem with a well-defined ambiguity set. The problem was further reformulated as a computationally tractable conic optimization problem. A numerical example is solved to illustrate the validity of the conic optimization problem. We believe that the result in this paper can provide a new way for applying multiple linear regression in practice.

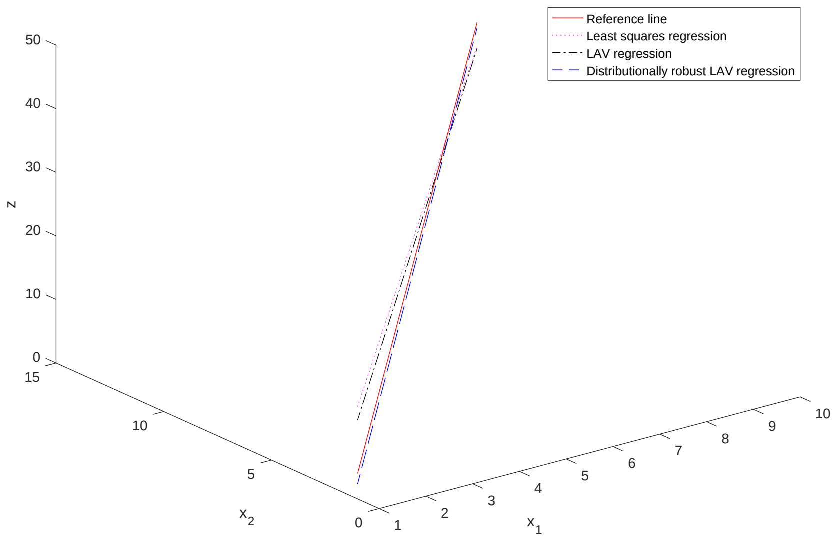

Fig. 2 The regression lines for numerical example.

Acknowledgements

This work was supported by the Natural Science Foundation of China (grants 11771008 and 71471103), the Natural Science Foundation of Shandong Province, China (grants ZR2017MA005 and ZR2015AL010), and the Australian Research Council (grants DP140100289 and DP160102819).

References

1. S. Weisberg, Applied Linear Regression, John Wiley & Sons, Inc., New Jersey, 2005
2. S.M. Stigler, The History of Statistics: The Measurement of Uncertainty Before 1900, The Belknap Press of Harvard University Press, Cambridge, 1986.
3. T.E. Dielman, Least absolute value estimation in regression models: An anotated bibliography, Communications in Statistic-Theory and Methods, 4 (1984), 513–541.
4. Y. Dodge, L_1-Statistical Analysis and Related Methods, North-Holland Publishing, Amsterdam, 1992.
5. Y. Dodge, W. Falconer, Statistical Data Analysis Based on the L_1-Norm and Related Methods, Barika Photography & Productions, New Bedford, 2002.
6. E. Damsleth, A.H. EI-Shaarawi, ARMA models with double-exponentially distributed noise, Journal of the Royal Statistical Society, Series B, 51 (1989), 61–69.
7. Z. Ding, C.W.J. Granger, Modeling volatility persistence of speculative returns: A new approach, Journal of Econometrics, 73 (1996), 185–215.
8. H.L. Harter, Nonuniqueness of least absolute values regression, Communications in Statistic-Theory and Methods, A6 (1977), 829–838.
9. A. Charnes, W.W. Cooper, R.O. Ferguson, Optimal estimation of executive compensation by linear programming, Management Science, 1 (1955), 138–151.
10. P. Bloomfield, W. Steiger, Least absolute deviations curve-fitting, SIAM Journal on Scientific and Statistical Computing, 1 (1980), 290–301.
11. Y. Zhang, Primal-dual intrior point approach for computing l_1-solutions, and l_∞-solutions of overdetermined linear systems, Journal of Optimization Theory and Applications, 77 (1993), 323–341.

12. H. Li, Solving least absolute value regression problem using modified goal programming techniques, Computers & Operations Research, 25 (1998), 1137–1143.
13. T.E. Dielman, Least absolute value regression: Recent contributions, Journal of Statistical Computation and Simulation, 75 (2005), 263–286.
14. L. EI Ghaoui, M. Oks, F. Oustry, Worst-case value-at-risk and robust portfolio optimization: A conic programming approach, Operations Research, 51 (2003), 543–556.
15. E. Delage, Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problem, Operations Research, 58 (2010), 596–612.
16. J. Goh, M. Sim, Distributionally robust optimization and its tractable approximations, Operations Research, 58 (2010), 902–917.
17. J. Ang, F. Meng, J. Sun, Two-stage stochastic linear programs with incomplete information on uncertainty, European Journal of Operational Research, 233 (2014), 16–22.
18. S. Gao, L. Kong, J. Sun, Robust two-stage stochastic linear programs with moment constraints, Optimization, 63 (2014), 829–837.
19. S. Gao, J. Sun, S. Wu, A semi-infinite programming approach to two-stage stochastic linear programs with high-order moment constraints, Optimization Letters, (2016), doi:10.1007/s11590-016-1095-4
20. S. Mehrotra, H. Zhang, Models and algorithms for distributionally robust least squares problems, Mathematical Programming, Series A, 146 (2014), 123–141.
21. J.R. Birge, F. Louveaux, Introduction to Stochastic Programming, Springer, New York, 1997.
22. D. Bertsimas, D.B. Brown, Constructing uncertainty sets for robust linear optimization, Operations Research, 57 (2009), 1483–1495.
23. W. Wiesemann, D. Kuhn, M. Sim, Distributionally robust convex optimization, Operations Research, 2 (2014), 1358–1376.
24. R.T. Rockafellar, Conjugate Duality and Optimization, SIAM, Philadelphia, 1974.
25. G. Wang, X.Q. Yang, T.C.E. Cheng, Generalized Levitin-Polyak well-posedness for generalized semi-infinite programs, Numerical Functional Analysis & Optimization, 34 (2013), 695–711.
26. A. Ben-Tal, L. El Ghaoui, A. Nemirovski, Robust Optimization, Princeton University Press, Princeton, 2009.
27. S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, UK, 2004.
28. M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, http://cvxr.com/cvx, 2017.