ORIGINAL - OR MODELING/CASE STUDY

A second-order cone programming based robust data envelopment analysis model for the new-energy vehicle industry

Chao Lu[1] · Jie Tao[2] · Qiuxian An[3] · Xiaodong Lai[4]

Abstract

The validity of performance evaluation is determined by, and therefore greatly influenced by, the accuracy of data set. To address such imprecise and negative data problems widely spread in the real world, this paper proposes a second-order cone based robust data envelopment analysis (SOCPR-DEA) model, which is more robust to data variety. Further, this new computational tractable model is applied to analyze 13 new-energy vehicle (NEV) manufacturers from China. The findings support that the SOCPR-DEA model could well mitigate the deficiency caused by data variety, and the evidence from Chinese NEV industry shows that a focus strategy is more likely to enhance a firm's efficiency especially at its emerging stage, and the efficiency is more sensitive with production cost than other factors such as research and development, sales income, earnings per share, and predicted income. In addition, this paper also gives some industrial implications and policy suggestions based on these interesting findings.

✉ Jie Tao
  taojie@usst.edu.cn

  Chao Lu
  06luchao@163.com

[1] School of Management, Shanghai University, No.99 Shangda Road, Baoshan District, Shanghai 200444, China

[2] Business School, University of Shanghai for Science and Technology, No. 516 Jungong Road, Shanghai 200093, China

[3] School of Business, North China Electric Power University, No. 2 Beinong Road, Beijing 100086, China

[4] School of Economic and Management, South China Normal University, Guangzhou 510631, China

1 Introduction

Since 2015, China has replaced US as the largest new-energy vehicle country, and the Civilian Auto Ownership of 2017 reached 1.8 million, taking more than 50% share of the whole world's amount. The rapid development of China's new-energy vehicle (NEV) industry comes from the pressure of national energy security and urban pollution control and emission reduction, as well as the desire from energy transformation and automobile industry revitalization (Fig. 1). However, this "corner overtaking" of China's NEV industry was widely criticized for its high relying on the support from both central and local governments (Lu et al. 2014), let alone that a large number of manufacturers defrauded financial subsidies as their important profit model (Lu et al. 2017). From the other side, there is still a long way for China's NEV industry to go before its large-scale industrialization and sustainable development, due to the immature key technologies (e.g. power battery, electric motor, and electronic controlling system), the insufficient charging infrastructure, and the imperfect user experience (Liu et al. 2018) . Although it is a consensus that firms should keep focusing on R&D to improve the quality of NEV products, most existing research mainly explored the entire NEV industry issues at the macro or medium level (Liu and Kokko 2013; Zhang and Qin 2018), while such studies on evaluating firm performances haven't yet been paid enough attention.

To carry out the performance evaluation of NEV firms, the classic methodology of Data Envelopment Analysis (DEA) firstly proposed by Charnes et al. (1978) , can be well employed for its strength in evaluating and measuring the relative efficiencies of a set of decision-making units (DMUs) that use multiple inputs to produce multiple outputs. Recently, the application of DEA to the energy industry especially the new-energy vehicle industry increased a lot, see Sueyoshi et al. (2017). When applying DEA models, how to improve the classical models to address the extensive uncertain issues in real-life environment is an increasingly important problem (Emrouznejad and Yang 2018). For example, the fuzzy DEA model (Wen 2015; Omrani et al. 2018), the stochastic DEA model (Cooper et al. 1996; Zhou et al. 2017; Chen et al. 2018), the chance-constrained DEA model (Chen and Zhu 2019). However, most of these models need to specify a probability distribution function for the error process, which may not be realistic (Talluri et al. 2006).

By considering data variety in real-world problems, a more robust DEA model should be not only translation invariant, but also immune to data uncertainty. For example, when the efficiencies of EV manufactures are under evaluation, variables such as net profit, earnings
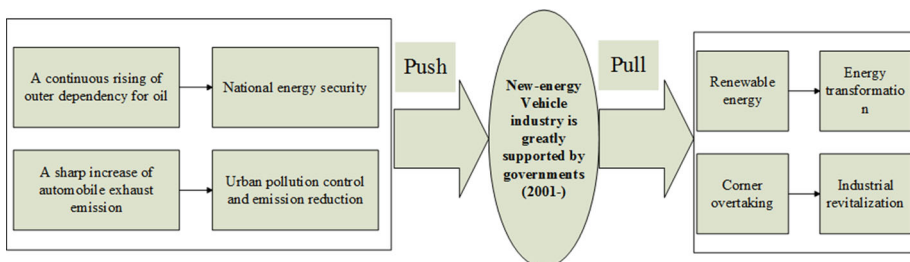
Fig. 1 The background of NEV industry's rapid development in China

per share (EPS) are usually considered. For some EV companies, these variables maybe zero value or even negative value. For example, EPS is the monetary value of earnings per outstanding share of common stock for a company. The formula of EPS is:

$$\text{EPS} = \frac{\text{Net Income} - \text{Preferred Dividends}}{\text{Average Common Shares}}.$$

Therefore the negative EPS means a company's Net Income is less than the Preferred Dividends, or sometimes, the Net Income is negative. This phenomenon is common for small companies, and sometimes it is a result of economic depression. From the financial statements of the Xiamen King Long United Automotive Industry Company Limited (KING LONG),[1] the EPS from 2013 to 2016 are 0.75, −0.17, 0.44 and 0.31. In order to address the negative data −0.17, the DEA model should be translation invariant. Moreover, if we want to evaluate the efficiency of EV manufacture companies in 2017, the value of some variables such as profit, can only be obtained by prediction which are not accurate, as some data are not available since the financial statement are usually published at the end of the year. Hence, in order to address the inaccurate data problem, the DEA model should be more robust.

Therefore, we integrate the technique of robust optimization to some particular classical DEA model with translation invariance property, so as to transform it into a second-order cone programming without any further assumptions on the type of data distribution. Furthermore, the Chinese NEV industry is a great example, to test the validation and verification of our improved DEA model, vice versa, the empirical analysis also gives some useful insights for the emerging industry.

This paper is structured as follows: following Sect. 1, Sect. 2 will firstly review the classical DEA models and multi-objective DEA (MDEA) model , and will then apply the robust optimization technique to MDEA model to develop the R-MDEA model. This is followed by integrating the R-MDEA model with the chance-constrained DEA model and proposing our new model named as the second-order cone based robust data envelopment analysis (SOCPR-DEA) model in the Sect. 3. Section 4 will make a practical evaluation on 13 new-energy vehicle manufacturers in China by applying the SOCPR-DEA model. Conclusions and directions for future research are discussed in the Sect. 5.

2 Multi-objective DEA model and robust optimization model

2.1 Multi-objective DEA model

The classical DEA models are expressed in model (1) by introducing three parameters to determine the type of the model. Specifically, in the case of $\sigma_1 = 0$, $\sigma_2 = 0$, $\sigma_3 = 0$, the model is C^2R, constant returns to scale. In the case of $\sigma_1 = 1$, $\sigma_2 = 0$, the model is BC^2, variable returns to scale. In the case of $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 0$, the model is FG, decreasing returns to scale. In the case of $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, the model is ST, increasing returns to scale (Yan and Wei 2011).

$$
\begin{aligned}
\max \ & \mu \mathbf{y}_0 - \sigma_1 \mu_0 \\
\text{s.t.} \ & \omega \mathbf{x}_j - \mu \mathbf{y}_j + \sigma_1 \mu_0 \geq 0, \ j = 1, \ldots, n, \\
& \omega \mathbf{x}_0 = 1, \\
& \omega \geq \mathbf{0}, \ \mu \geq \mathbf{0}, \ \sigma_1 \sigma_2 (-1)^{\sigma_3} \mu_0 \geq 0.
\end{aligned}
\tag{1}
$$

[1] http://money.finance.sina.com.cn/corp/go.php/vFD_FinancialGuideLine/stockid/600686/ctrl/2013/displaytype/4.phtml.

The dual form and the production possibility set of model (1) is expressed in model (2) and model (3). Certainly they satisfy the basic optimization theories, e.g., the strong duality theorem, etc.

$$
\begin{aligned}
\min\ & \theta \\
\text{s.t.}\ & \sum_{j=1}^{n} \mathbf{x}_j \lambda_j \leq \theta \mathbf{x}_0, \\
& \sum_{j=1}^{n} \mathbf{y}_j \lambda_j \geq \mathbf{y}_0, \\
& \sigma_1 \left(\sum_{j=1}^{n} \lambda_j + \sigma_2 (-1)^{\sigma_3} \lambda_{n+1} \right) = \sigma_1, \\
& \lambda_j \geq 0,\ j = 1, \ldots, n, n+1.
\end{aligned}
\tag{2}
$$

$$
T = \left\{ (\mathbf{x}, \mathbf{y}) :
\begin{array}{c}
\sum_{j=1}^{n} \mathbf{x}_j \lambda_j \leq \mathbf{x},\ \sum_{j=1}^{n} \mathbf{y}_j \lambda_j \geq \mathbf{y}, \\
\sigma_1 \left(\sum_{j=1}^{n} \lambda_j + \sigma_2 (-1)^{\sigma_3} \lambda_{n+1} \right) = \sigma_1, \\
\lambda_j \geq 0,\ j = 1, \ldots, n, n+1.
\end{array}
\right\}
\tag{3}
$$

However, the classical DEA model can only maximize (minimize) the outputs (inputs) when the inputs (outputs) are fixed while neglecting the differences of inputs (outputs) among different DMUs. Strictly speaking, the efficiency measured in classical DEA models is the single criterion efficiency which only considers the DMU that is being evaluated. To describe the efficiency more objectively, Chambers et al. (1998) proposed the directional distance function model (DDF model), and explained the duality relationship between the DDF model and the profit function .

$$
\begin{aligned}
\max\ & \beta \\
\text{s.t.}\ & \sum_{j=1}^{n} \lambda_j x_{ij} \leq x_{io} - \beta g_i^x,\ i = 1, \ldots, m, \\
& \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro} + \beta g_r^y,\ r = 1, \ldots, s, \\
& \sum_{j=1}^{n} \lambda_j = 1, \\
& \lambda_j \geq 0,\ j = 1, \ldots, n,
\end{aligned}
\tag{4}
$$

where $G = (g^x, g^y) \in \mathbb{R}_+^m \times \mathbb{R}_+^s$ is the non-negative directional input–output vector. Model (4) is the general form of the DDF model. Using duality theory, Chambers et al. (1998) show that the DDF is the lower bound of the profit inefficiency measure (the Mahler inequality), and in this sense, the profit inefficiency can be decomposed into the technical inefficiency part and the allocative inefficiency part. Based on the theory work of the DDF model, a number of profit inefficiency measures have been investigated (Aparicio et al. 2014). Further, in comparison with the radial direction models (e.g., CCR, BCC), the DDF model can be classified into the slacks-based category. The slacks-based measure model (SBM) was first proposed by Tone (2001), and further investigated by Tone and Tsutsui (2010) and You and Jie (2016).

$$\min \rho = \frac{1 - \frac{1}{m} \sum\limits_{i=1}^{m} \frac{s_i^-}{x_{io}}}{1 - \frac{1}{s} \sum\limits_{r=1}^{s} \frac{s_r^+}{y_{ro}}}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} \lambda_j x_{ij} = x_{io} - s_i^-, i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} = y_{ro} + s_r^+, r = 1, \ldots, s, \quad (5)$$

$$\sum_{j=1}^{n} \lambda_j = 1,$$

$$\lambda_j \geq 0, j = 1, \ldots, n, s_i^-, s_r^+ \geq 0.$$

The main virtues of the SBM model are units invariant, monotone decreasing with respect to input excess and output shortfall, and the efficiency value measured by the SBM model is between nil and unity. Furthermore, in contrast to the classical CCR model, the dual side of the SBM model can also be interpreted as profit maximization.

Based on the previous work of the DDF model and the SBM model, Estellita Lins et al. (2004) proposed the Multi-Objective DEA (MDEA) model.

$$\max \quad \omega = \frac{1}{s} \sum_{r=1}^{s} \eta_s - \frac{1}{m} \sum_{i=1}^{m} \theta_i$$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} \lambda_j - \theta_i x_{i0} \leq 0, i = 1, \ldots, m,$$

$$- \sum_{j=1}^{n} y_{rj} \lambda_j + \eta_r y_{r0} \leq 0, r = 1, \ldots, s, \quad (6)$$

$$0 \leq \theta_i \leq 1, \eta_r \geq 1,$$

$$\sum_{j=1}^{n} \lambda_j = 1,$$

$$\lambda_j \geq 0, j = 1, \ldots, n.$$

Theorem 1 *DMU_{j0} is MDEA efficient if and only if $\omega = 0$, $\theta_i = 1$, $\eta_r = 1$, $i = 1, \ldots, m$, $r = 1, \ldots, s$ (Estellita Lins et al. 2004).*

As Theorem 1 indicates, if DMU_{j0} is MDEA efficient, the value of the objective function is nil. Conversely, because $0 \leq \theta_i \leq 1, \eta_r \geq 1$, we may conclude that as ω increases, the efficiency of a given DMU decreases. The advantage of the MDEA model over the classical DEA models in detail could be referred in Estellita Lins et al. (2004).

It is noted that the MDEA model can be seen as a extension of the DDF model and the SBM model and it inherits all the virtues of the two models. First, if we set $g_i^x = x_{io}, g_r^y = y_{ro}$ in model (4), the constraints in model (6) can be seen as the combination of the constraints in model (4) and model (5). In the DDF model, whatever g_i^x, g_r^y might be, the constraints in model (4) were β related. Obviously, both the SBM model and the MDEA overcame this little flaw. Second, the efficiency value measured in the DDF model was also β related, while the efficiency value of efficient DMUs is nil and unity in the MDEA model and the SBM model respectively. Third, the structure of the MDEA model is simpler than that of the SBM model. Moreover the MDEA model possesses the translation invariance property so as to address the negative and zero data problem. Therefore, we here use the MDEA model as our basic model for further analysis.

2.2 Robust multi-objective DEA model

Model (4) shows the case where the inputs and outputs are deterministic. However, when there exists some uncertain factors in the inputs and outputs set, model (4) is not useful. Therefore in this section we develop a robust multi-objective DEA model to address the nondeterministic data problem.

Robust optimization is one of the main methods of dealing with uncertain program, which is first proposed by Soyster (1973). The arise of robust optimization is due to the drawbacks of the stochastic program, which are often hard or impossible to solve directly. Solutions of robust optimization often perform well when analyzed under stochastic models of parameter variation. The development of the robust optimization can be divided into two stages. One is based on scenarios, which is proposed by Mulvey et al. (1995), and later researched by Laguna (1998), Malcolm and Zenios (1994) and Vassiadou and Zenios (1996). The main drawback of the scenario based robust optimization programming is that it changes the structure of the original deterministic program. The other method is based on the property of convex sets (Yu and Li 2000; Leung et al. 2002).

Besides, Ben-Tal and Nemirovski (1998) have developed the basic framework of robust optimization by defining the uncertain set with the intersection of ellipsoids. El Ghaoui and Lebret (1997) have extended the Ben-Tal and Nemirovski's method to the least-squares problems with uncertain data. Bertsimas and Sim (2004) has introduced a Γ_i-based method by transforming the uncertain problem into a linear program. The main difference between these set based robust models is due to their disparate uncertain sets. Ben-Tal and Nemirovski (1998) use ellipsoidal uncertainty sets while the uncertainty set proposed by Bertsimas and Sim (2004) is a polyhedron that encode a budget of uncertainty in terms of cardinality constraints.

Because D. Betasimas's robust optimization model is a linear optimization model, more research on robust DEA model followed from D. Betasimas' idea (Shokouhi et al. 2010; Sadjadi and Omrani 2008; Gharakhani et al. 2011). However, there are many advantages when the uncertainty set is "ellipsoid" (Ben-Tal and Nemirovski 1999).

Furthermore, a polyhedron is a special case of the ellipsoid family, and the corresponding robust program could be transformed into a linear program when the ellipsoid uncertainty degenerates to a polyhedron (Ben-Tal and Nemirovski 1999). Based on the MDEA model, we propose a robust multi-objective DEA (R-MDEA) model by introducing the Ben-Tal and Nemirovski's robust optimization idea into the classical DEA models.

Considering the uncertain inputs and outputs set, we could impose constraints on the model such as $x_{ij}, y_{rj} \in \mho$, where $\mho$ is the uncertain set. The uncertainty set $\mho$ can be constructed as follows

$$\mho = \left\{ [A; \mathbf{0}] = [A^0; \mathbf{0}] + \sum_{l=1}^{L} \zeta [A^l; \mathbf{0}] | \zeta \in Z^L \right\},$$

where

$$A = \begin{bmatrix} x_{11} & \cdots & x_{1n} & -x_{10} & \cdots & 0 & 0 & \cdots & 0 \\ x_{21} & \cdots & x_{2n} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_{m1} & \cdots & x_{mn} & 0 & \cdots & -x_{m0} & 0 & \cdots & 0 \\ -y_{11} & \cdots & -y_{1n} & 0 & \cdots & 0 & y_{10} & \cdots & 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ -y_{s1} & \cdots & -y_{sn} & 0 & \cdots & 0 & 0 & \cdots & y_{s0} \end{bmatrix}.$$

Let A^0 be the nominal value matrix of A, A^l be the basic shift, $Z = \{\zeta | \zeta \in Z^l\}$ be the perturbation set. Thus, the MDEA model containing the uncertain set can be transformed into a semi-infinite program. However, the form of the uncertain set can be constructed by various perturbation sets Z; therefore it is possible only to consider Z for the construction of a complete uncertain set. Consequently, the MDEA model containing the uncertain set can be expressed as follows:

$$
\begin{aligned}
\max \ \omega &= \frac{1}{s} \sum_{r=1}^{s} \eta_s - \frac{1}{m} \sum_{i=1}^{m} \theta_i \\
\text{s.t.} \ \ & \sum_{j=1}^{n} x_{ij} \lambda_j - \theta_i x_{i0} \le 0, i = 1, \dots, m, \\
& -\sum_{j=1}^{n} y_{rj} \lambda_j + \eta_r y_{r0} \le 0, r = 1, \dots, s, \\
& 0 \le \theta_i \le 1, \eta_r \ge 1, \\
& \sum_{j=1}^{n} \lambda_j = 1, \\
& \lambda_j \ge 0, j = 1, \dots, n, \\
& \mathbf{x}, \mathbf{y} \in \mho, \\
& \mho = \left\{ [A; \mathbf{0}] = [A^0; \mathbf{0}] + \sum_{l=1}^{L} \zeta_l [A^l; \mathbf{0}] | \zeta \in Z^L \right\}.
\end{aligned}
\tag{7}
$$

In general, setting

$$
\begin{aligned}
\mho &= \left\{ [A; \mathbf{0}] = [A^0; \mathbf{0}] + \sum_{l=1}^{L} \zeta_l [A^l; \mathbf{0}] | \zeta \in Z^L \right\} \\
Z^L &= \left\{ \zeta \in \mathbf{R}^L | \exists \mathbf{u} \in \mathbf{R}^K \text{ s.t. } P\zeta + Q\mathbf{u} + \mathbf{p} \in K \right\},
\end{aligned}
\tag{8}
$$

where $K \in \mathbf{R}^N$ is a closed convex cone. Also note that P and Q are matrices with coefficients that are known.

Proposition 1 *Any ellipsoidal uncertainty $\mho = \{A = \Pi(\zeta) | \|P\zeta\|_2 \le 1\}$ could be represented by* (8).

Proof Set $[A^0; \mathbf{0}] = P^0$, $[A^l; \mathbf{0}] = P^l$, then the original uncertainty set $\mho$ could be expressed as $\mho = \{A = P^0 + \sum_{l=1}^{L} \zeta_l P^l | \zeta \in Z^L\}$, where $Z^L = \{\zeta \in \mathbf{R}^L | \exists \mathbf{u} \in \mathbf{R}^K \text{ s.t. } P\zeta + Q\mathbf{u} + \mathbf{p} \in K\}$. We analyze $\mho$ in two steps. First let us denote by $\mathbf{r}_i^j$ ith row of A^j, $j = 0, \dots, L$, and let R_i be the matrix with the columns $\mathbf{r}_i^1, \dots, \mathbf{r}_i^L$, so that the ith row of $\Pi(\zeta)$ is exactly $\mathbf{r}_i^0 + R_i \zeta$. Second, let us set some parameters in Z^L be $Q = 0$, $\mathbf{p} = [0, \dots, 0, 1]^T$, and K is a Lorentz cone. Therefore, any ζ in Z^L could be transformed as $\|P\zeta\|_2 \le 1$. In conclusion, we finally get $\mho = \prod_{l=1}^{L} \mho_l$ is constraint-wise uncertainty with ellipsoid $\mho_l$: $\mho_l = \{A | A^T \mathbf{e}_l = \mathbf{r}_l + R_l \zeta_l, \|\zeta\|_2 \le 1, l = 1, \dots, L\}$. The Proposition follows. □

Theorem 2 *If $Z = \left\{ \zeta \in \mathbf{R}^L | \exists \mathbf{u} \in \mathbf{R}^K \text{ s.t. } P\zeta + Q\mathbf{u} + \mathbf{p} \in K \right\}$, here $K \in \mathbf{R}^N$, the MDEA model* (7) *containing the uncertain set can be converted from a semi-infinite program into a cone-constrained MDEA model (R-MDEA model):*

$$\max \omega = \frac{1}{s} \sum_{r=1}^{s} \eta_s - \frac{1}{m} \sum_{i=1}^{m} \theta_i$$

$$s.t. \quad \mathbf{p}^T \alpha + \sum_{j=1}^{n} x_{ij}^0 \lambda_j - \theta_i x_{i0}^0 \leq 0, i = 1, \ldots, m,$$

$$\mathbf{p}^T \alpha - \sum_{j=1}^{n} y_{rj}^0 \lambda_j + \eta_r y_{r0}^0 \leq 0, r = 1, \ldots, s,$$

$$Q^T \alpha = \mathbf{0},$$

$$(P^T \alpha)_l + \sum_{j=1}^{n} x_{ij}^l \lambda_j - x_{x0}^l \theta_i \leq 0, i = 1, \ldots, m, \qquad (9)$$

$$(P^T \alpha)_l - \sum_{j=1}^{n} y_{rj}^l \lambda_j + y_{r0}^l \eta_r \leq 0,$$

$$\alpha \in K_* = \left\{ \mathbf{y} | \mathbf{y}^T \mathbf{z} \geq 0, \forall \mathbf{z} \in K \right\},$$

$$0 \leq \theta_i \leq 1, \eta_r \geq 1,$$

$$\sum_{j=1}^{n} \lambda_j = 1,$$

$$\lambda_j \geq 0, j = 1, \ldots, n.$$

Proof Setting $\beta = (\lambda_1, \ldots, \lambda_n, \theta_1, \ldots, \theta_m, \eta_1, \ldots, \eta_s)$ as the feasible solution of model (7), then $\sup_{\zeta \in Z} \left\{ (A^0 \beta) + \sum_{l=1}^{L} \zeta_l (A^l \beta) \right\} \leq 0$. Further, we could get

$$\max \left\{ (A^l \beta)^T \zeta | P \zeta + Q \mathbf{u} + \mathbf{p} \in K \right\} \leq -A^0 \beta. \qquad (10)$$

That is, β is a feasible solution of model (2) only when (10) is established. Applying the dual cone theorem (Boyd and Vandenberghe 2004), the dual cone program of (10) can be written as

$$\min_{\alpha} \left\{ \mathbf{p}^T \alpha | Q^T \alpha = -A^l \beta, \alpha \in K_* \right\} \leq -A^0 \beta. \qquad (11)$$

The inequality (10) is established if and only if there exists an α that satisfies (11). In other words, $\boldsymbol{\beta}$ is a feasible solution of model (7) if and only if there exists an α that satisfies (11). Then, we find that

$$\mathbf{p}^T \alpha + A^0 \boldsymbol{\beta} = \mathbf{p}^T \alpha + \sum_{j=1}^{n} x_{ij}^0 \lambda_j - \theta_i x_{i0}^0 \leq 0, i = 1, \ldots, m,$$

$$\mathbf{p}^T \alpha - \sum_{j=1}^{n} y_{rj}^0 \lambda_j + \eta_r y_{r0}^0 \leq 0, r = 1, \ldots, s,$$

$$Q^T \alpha = \mathbf{0}, \alpha \in K_*,$$

$$P^T \alpha + A^l \beta = (P^T \alpha)_l + \sum_{j=1}^{n} x_{ij}^l \lambda_j - x_{x0}^l \theta_i \leq 0, i = 1, \ldots, m,$$

$$(P^T \alpha)_l - \sum_{j=1}^{n} y_{rj}^l \lambda_j + y_{r0}^l \eta_r \leq 0.$$

The theorem follows. Therefore, we finally construct the R-MDEA model as (9).   □

3 The design of the SOCPR-DEA model

In this section, we integrate the R-MDEA model in the previous section and techniques of the chance constrained optimization, to propose the SOCPR-DEA model.

Suppose that there are n decision-making units : $DMU_1, DMU_2, \ldots, DMU_n$, coupled with m random input variables, $\tilde{\mathbf{X}}_j = (\tilde{x}_{1j}, \ldots, \tilde{x}_{mj})$, $j = 1, \ldots, n$, s represents the number of random output variables, $\tilde{\mathbf{Y}}_j = (\tilde{y}_{1j}, \ldots, \tilde{y}_{sj})$, $j = 1, \ldots, n$. Setting the confidence level as $1 - \alpha, 0 \leq \alpha \leq 1$, the multi-objective chance-constrained model can be written as

$$
\begin{aligned}
\min \quad & (\theta_1, \cdots, \theta_m, -\eta_1, \ldots, -\eta_s) \\
\text{s.t.} \quad & P\left(\sum_{j=1}^{n} \tilde{x}_{ij}\lambda_j \leq \theta_i \tilde{x}_{i0}\right) \geq 1 - \alpha, i = 1, \ldots, m, \\
& P\left(\sum_{j=1}^{n} \tilde{y}_{rj}\lambda_j \geq \eta_r \tilde{y}_{r0}\right) \geq 1 - \alpha, r = 1, \ldots, s, \\
& \eta_r \geq 1, 0 \leq \theta_i \leq 1, i = 1, \cdots, m, r = 1, \ldots, s, \\
& \sum_{j=1}^{n} \lambda_j = 1, \lambda_j \geq 0, j = 1, \ldots, n.
\end{aligned}
\tag{12}
$$

Because model (12) is a multi-objective program, we could define $\max \omega = \frac{1}{s}\sum_{r=1}^{s} \eta_r - \frac{1}{m}\sum_{i=1}^{m} \theta_i$ as a substitution to the objective function of (12). Furthermore, DMU_{j0} is $(1-\alpha) * 100\%$ efficient if and only if $\omega = 0$. The previous approach for solving the chance-constrained MDEA model involves specifying a probability distribution function (e.g., normal) for the error process. However, because the probability distribution of these random variables cannot always be known in advance, the solution will deviate from the actual result. Assume that

$$
\begin{bmatrix}
\tilde{x}_{11} & \tilde{x}_{12} & \cdots & \tilde{x}_{1n} \\
\tilde{x}_{21} & \tilde{x}_{22} & \cdots & \tilde{x}_{2n} \\
\cdots & \cdots & \cdots & \cdots \\
\tilde{x}_{m1} & \tilde{x}_{m2} & \cdots & \tilde{x}_{mn} \\
\tilde{y}_{11} & \tilde{y}_{12} & \cdots & \tilde{y}_{1n} \\
\tilde{y}_{21} & \tilde{y}_{22} & \cdots & \tilde{y}_{2n} \\
\cdots & \cdots & \cdots & \cdots \\
\tilde{y}_{s1} & \tilde{y}_{s2} & \cdots & \tilde{y}_{sn}
\end{bmatrix}
=
\begin{bmatrix}
x_{11}^0 & x_{12}^0 & \cdots & x_{1n}^0 \\
x_{21}^0 & x_{22}^0 & \cdots & x_{2n}^0 \\
\cdots & \cdots & \cdots & \cdots \\
x_{m1}^0 & x_{m2}^0 & \cdots & x_{mn}^0 \\
y_{11}^0 & y_{12}^0 & \cdots & y_{1n}^0 \\
y_{21}^0 & y_{22}^0 & \cdots & y_{2n}^0 \\
\cdots & \cdots & \cdots & \cdots \\
y_{s1}^0 & y_{s2}^0 & \cdots & y_{sn}^0
\end{bmatrix}
+
\sum_{l=1}^{l} \zeta_l
\begin{bmatrix}
x_{11}^l & x_{12}^l & \cdots & x_{1n}^l \\
x_{21}^l & x_{22}^l & \cdots & x_{2n}^l \\
\cdots & \cdots & \cdots & \cdots \\
x_{m1}^l & x_{m2}^l & \cdots & x_{mn}^l \\
y_{11}^l & y_{12}^l & \cdots & y_{1n}^l \\
y_{21}^l & y_{22}^l & \cdots & y_{2n}^l \\
\cdots & \cdots & \cdots & \cdots \\
y_{s1}^l & y_{s2}^l & \cdots & y_{sn}^l
\end{bmatrix},
$$

where $\zeta \in Z^L$ and

$$
\begin{bmatrix}
x_{11}^0 & x_{12}^0 & \cdots & x_{1n}^0 \\
x_{21}^0 & x_{22}^0 & \cdots & x_{2n}^0 \\
\cdots & \cdots & \cdots & \cdots \\
x_{m1}^0 & x_{m2}^0 & \cdots & x_{mn}^0 \\
y_{11}^0 & y_{12}^0 & \cdots & y_{1n}^0 \\
y_{21}^0 & y_{22}^0 & \cdots & y_{2n}^0 \\
\cdots & \cdots & \cdots & \cdots \\
y_{s1}^0 & y_{s2}^0 & \cdots & y_{sn}^0
\end{bmatrix}
$$

is the nominal value matrix of random input and output variables.

$$
\begin{bmatrix}
x_{11}^l & x_{12}^l & \cdots & x_{1n}^l \\
x_{21}^l & x_{22}^l & \cdots & x_{2n}^l \\
x_{m1}^l & x_{m2}^l & \cdots & x_{mn}^l \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
y_{11}^l & y_{12}^l & \cdots & y_{1n}^l \\
y_{21}^l & y_{22}^l & \cdots & y_{2n}^l \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
y_{s1}^l & y_{s2}^l & \cdots & y_{sn}^l
\end{bmatrix}
$$

is the basic shift matrix, and $Z = \{\zeta \mid \zeta \in Z^L\}$ is the perturbation set.

Accordingly, the first chance constraint of the original problem can be expressed as

$$
\sum_{j=1}^{n} \tilde{x}_{ij}\lambda_j - \theta_i \tilde{x}_{i0}
$$

$$
= \sum_{j=1}^{n} \left(x_{ij}^0 + \sum_{l=1}^{L} \zeta_l x_{ij}^l \right) \lambda_j - \theta_i \left(x_{i0}^0 + \sum_{l=1}^{L} \zeta_l x_{i0}^l \right)
$$

$$
= \left(\sum_{j=1}^{n} x_{i0}^0 \lambda_j - \theta_i x_{i0}^0 \right) + \sum_{l=1}^{L} \zeta_l \left(\sum_{j=1}^{n} x_{ij}^l \lambda_j - \theta_i x_{i0}^l \right) \le 0, i = 1, \ldots, m. \quad (13)
$$

Similarly, the second chance constraint of the original problem can be expressed as

$$
\left(-\sum_{j=1}^{n} y_{rj}^0 \lambda_j + \eta_r y_{r0}^0 \right) + \sum_{l=1}^{L} \zeta_l \left(\sum_{j=1}^{n} y_{rj}^l \lambda_j + \eta_r y_{r0}^l \right) \le 0, r = 1, \ldots, s. \quad (14)
$$

Therefore, setting the confidence level as $(1-\alpha)$, $0 \le \alpha \le 1$, model (12) can be converted into

$$
\max \omega = \frac{1}{s} \sum_{r=1}^{s} \eta_s - \frac{1}{m} \sum_{i=1}^{m} \theta_i
$$

$$
\text{s.t. } P\left\{ \left(\sum_{j=1}^{n} x_{i0}^0 \lambda_j - \theta_i x_{i0}^0 \right) + \sum_{l=1}^{L} \zeta_l \left(\sum_{j=1}^{n} x_{ij}^l \lambda_j - \theta_i x_{i0}^l \right) \le 0 \right\} \ge 1-\alpha, \quad i=1,\ldots,m,
$$

$$
P\left\{ \left(-\sum_{j=1}^{n} y_{rj}^0 \lambda_j + \eta_r y_{r0}^0 \right) + \sum_{l=1}^{L} \zeta_l \left(\sum_{j=1}^{n} y_{rj}^l \lambda_j + \eta_r y_{r0}^l \right) \le 0 \right\} \ge 1-\alpha, \quad r=1,\ldots,s, \quad (15)
$$

$$
\eta_r \ge 1, 0 \le \theta_i \le 1, i = 1, \ldots, m, r = 1, \ldots, s,
$$

$$
\sum_{j=1}^{n} \lambda_j = 1, \lambda_j \ge 0, j = 1, \ldots, n.
$$

Assume that $\{\zeta_l\}$ are independent variables that satisfy $E(\zeta_l) = 0$, $\|\zeta_l\| \le 1, l = 1, \ldots, l$. It is noted that for any $\zeta_l \in [a, b]$, this constraint can be mapped to the interval $[-1, 1]$ by first transforming it into $[\frac{a-b}{2}, \frac{b-a}{2}]$ and later scaling the result by $\frac{b-a}{2}$.

Setting

$$\beta_1 = \sum_{l=1}^{L} \zeta_l \left(\sum_{j=1}^{n} x_{ij}^l \lambda_j - \theta_i x_{i0}^l \right) \leq \theta_i x_{i0}^0 - \sum_{j=1}^{n} x_{i0}^0 \lambda_j,$$

$$\beta_2 = \sum_{l=1}^{L} \zeta_l \left(-\sum_{j=1}^{n} y_{rj}^l \lambda_j + \eta_r y_{r0}^l \right) \leq -\eta_r y_{r0}^0 + \sum_{j=1}^{n} y_{r0}^0 \lambda_j,$$

we get

$$\text{STD}(\beta_1) = \sqrt{\sum_{l=1}^{L} \left(\sum_{j=1}^{n} x_{ij}^l \lambda_j - \theta_i x_{i0}^l \right)^2 E(\zeta_l^2)} \leq \sqrt{\sum_{l=1}^{L} \left(\sum_{j=1}^{n} x_{ij}^l \lambda_j - \theta_i x_{i0}^l \right)^2},$$

$$\text{STD}(\beta_2) = \sqrt{\sum_{l=1}^{L} \left(-\sum_{j=1}^{n} y_{rj}^l \lambda_j + \eta_r y_{r0}^l \right)^2 E(\zeta_l^2)} \leq \sqrt{\sum_{l=1}^{L} \left(-\sum_{j=1}^{n} y_{rj}^l \lambda_j + \eta_r y_{r0}^l \right)^2}.$$

Therefore, the chance constraints become

$$P \left\{ \beta_1 \leq \sqrt{\sum_{l=1}^{L} \left(\sum_{j=1}^{n} x_{ij}^l \lambda_j - \theta_i x_{i0}^l \right)^2} \right\} \geq 1 - \alpha,$$

$$P \left\{ \beta_2 \leq \sqrt{\sum_{l=1}^{L} \left(-\sum_{j=1}^{n} y_{rj}^l \lambda_j + \eta_r y_{r0}^l \right)^2} \right\} \geq 1 - \alpha,$$

which may be satisfied by setting the threshold Ω as shown below.

$$\Omega \sqrt{\sum_{l=1}^{L} \left(\sum_{j=1}^{n} x_{ij}^l \lambda_j - \theta_i x_{i0}^l \right)^2} \leq \theta_i x_{i0}^0 - \sum_{j=1}^{n} x_{i0}^0 \lambda_j,$$

$$\Omega \sqrt{\sum_{l=1}^{L} \left(-\sum_{j=1}^{n} y_{rj}^l \lambda_j + \eta_r y_{r0}^l \right)^2} \leq -\eta_r y_{r0}^0 + \sum_{j=1}^{n} y_{rj}^0 \lambda_j.$$

It is noted that as Ω increases, the possibility of the following variable increases:

$$\beta_1 \leq \theta_i x_{i0}^0 - \sum_{j=1}^{n} x_{i0}^0 \lambda_j,$$

$$\beta_2 \leq -\eta_r y_{r0}^0 + \sum_{j=1}^{n} y_{rj}^0 \lambda_j.$$

Lemma 1 *Setting $z_l, l = 1, \ldots, L$ as known coefficients, $\{\zeta_l\}, l = 1, \ldots, L$ are independent variables that satisfy $E(\{\zeta_l\}) = 0, |\zeta_l|$, then for any $\Omega \geq 0$, we have* $P\left\{ \zeta | \sum_{l=1}^{L} z_l \zeta_l > \Omega \sqrt{\sum_{l=1}^{L} z_l^2} \right\} \leq \exp\left(-\frac{\Omega^2}{2} \right)$ (Ben-Tal et al. 2009).

Corollary 1 *Replacing the constraints in model* (15) *with*

$$\Omega \sqrt{\sum_{l=1}^{L} \left(\sum_{j=1}^{n} x_{ij}^{l} \lambda_j - \theta_i x_{i0}^{l} \right)^2} \leq \theta_i x_{i0}^{0} - \sum_{j=1}^{n} x_{ij}^{0} \lambda_j,$$

$$\Omega \sqrt{\sum_{l=1}^{L} \left(-\sum_{j=1}^{n} y_{rj}^{l} \lambda_j + \eta_r y_{r0}^{l} \right)^2} \leq -\eta_r y_{r0}^{0} + \sum_{j=1}^{n} y_{rj}^{0} \lambda_j,$$

we can guarantee that

$$P \left\{ \left(\sum_{j=1}^{n} x_{i0}^{0} \lambda_j - \theta_i x_{i0}^{0} \right) + \sum_{l=1}^{L} \zeta_l \left(\sum_{j=1}^{n} x_{ij}^{l} \lambda_j - \theta_i x_{i0}^{l} \right) > 0 \right\} \leq \mathbf{exp} \left(-\frac{\Omega^2}{2} \right),$$

$$P \left\{ \left(-\sum_{j=1}^{n} y_{r0}^{0} \lambda_j + \eta_r y_{r0}^{0} \right) + \sum_{l=1}^{L} \zeta_l \left(\sum_{j=1}^{n} y_{rj}^{l} \lambda_j + \eta_r y_{r0}^{l} \right) > 0 \right\} \leq \mathbf{exp} \left(-\frac{\Omega^2}{2} \right).$$

In particular, when $\Omega > \sqrt{2 \ln \left(\frac{1}{\alpha} \right)}$*, the violating probability is less than* α.

Now, we consider a common perturbation set $Z = \{ \zeta \in R^L | |\zeta_l| < 1, l = 1, \ldots, L, \ \|\zeta\|_2 < \Omega \}$ (Ben-Tal et al. 2009). It is clear that Z is an intersection of the unit box with a ball of radius Ω that is centered at the origin. We have two reasons to choose Z as a perturbation set: One reason is that the unit box limits each component of the perturbation vector ζ between $[-1, 1]$, which ensures that each uncertain data a_{ij} is confined within the interval $[a_{ij}^{0} - \hat{a}_{ij}, a_{ij}^{0} + \hat{a}_{ij}]$. The other reason is that the ball of radius Ω centered at the origin guarantees an immunization of $1 - \mathbf{exp} \left(-\frac{\Omega^2}{2} \right) \times 100\%$.

Further, Ben-Tal et al. (2009) pointed out that the constraints of the robust program:

$$\min \left\{ \mathbf{c}^T \mathbf{x} + d | P(A\mathbf{x} \leq \mathbf{b}) \leq \alpha, (A, \mathbf{b}) \in \mho \right\},$$

$$\mho = \left\{ [\mathbf{a}_i; b_i] = [\mathbf{a}_i^0; b_i^0] + \sum_{l=1}^{L} \zeta_l [\mathbf{a}_i^l; b_i^l] | \zeta \in Z \right\},$$

can be converted to following secondary cone constraint: $z_l + \omega_l = b^l - (\alpha^l)^T \mathbf{x}, l = 1, \ldots, L, \sum_{l=1}^{L} |z_l| + \Omega \sqrt{\sum_{l=1}^{L} \omega_l^2} \leq b^0 - (\alpha^0)^T \mathbf{x}$. Moreover, the probability of a feasible solution under the new constraints is also feasible to the original program, and the probability is no lower than $1 - \mathbf{exp} \left(-\frac{\Omega^2}{2} \right)$. Accordingly, the multi-objective chance-constrained DEA

model can ultimately be converted to

$$\max \omega = \frac{1}{s} \sum_{r=1}^{s} \eta_s - \frac{1}{m} \sum_{i=1}^{m} \theta_i$$

$$\text{s.t.} \quad z_{li} + \omega_{li} = - \sum_{j=1}^{n} \lambda_j x_{ij}^l + \theta_i x_{i0}^l, \quad i = 1, \ldots, m,$$

$$z_{lr} + \omega_{lr} = \sum_{j=1}^{n} \lambda_j y_{rj}^l - \eta_r y_{r0}^l, \quad r = 1, \cdots, s,$$

$$\sum_{l=1}^{L} |z_{li}| + \Omega \sqrt{\sum_{l=1}^{L} \omega_{li}^2} \leq - \sum_{j=1}^{n} \lambda_j x_{ij}^0 + \theta_i x_{i0}^0, \quad i = 1, \ldots, m, \qquad (16)$$

$$\sum_{l=1}^{L} |z_{lr}| + \Omega \sqrt{\sum_{l=1}^{L} \omega_{lr}^2} \leq \sum_{j=1}^{n} \lambda_j y_{rj}^0 - \eta_r y_{r0}^0, \quad r = 1, \ldots, s,$$

$$\eta_r \geq 1, 0 \leq \theta_i \leq 1, \quad i = 1, \cdots, m, r = 1, \ldots, s,$$

$$\sum_{j=1}^{n} \lambda_j = 1, \lambda_j \geq 0, \quad j = 1, \ldots, n.$$

Similar to the definition of an efficient DMU in an MDEA model, if DMU$_{j0}$ is efficient in MDEA, the objective function value is 0, which implies that $\omega = 0, \theta_i = 1, \eta_r = 1, i = 1, \ldots, m, r = 1, \ldots, s$.

4 A case study of Chinese NEV industry

In this section, we will apply the new SOCPR-DEA model to carry out an efficiency evaluation by using 13 manufacturers from Chinese NEV industry as the case. The case study will not only test the validation and verification of SOCPR-DEA model, but also help to identify the main influence factors that concern the development of Chinese NEV industry.

As illustrated in the first column of Table 1, 13 Chinese new-energy vehicle manufactures as the DMUs to carry out the efficiency evaluation. The 13 auto manufacturers are the main NEV firms in China which contribute most authorized NEV products approved by the Ministry of Industry and Information Technology, covering both passenger cars and buses. With reference to the criteria of variables selection in Lee et al. (2012) and taking the suitability of SOCPR-DEA model into consideration, we select two inputs as research and development expenses (hereafter referred to as R&D) and production cost (hereafter referred to as PC), as well as three outputs as sales income (hereafter referred to as SI), earnings per share (hereafter referred to as EPS) and the predicted income of 2018 (hereafter referred to as PI), and PI is an ideal index that reflects the development prospects for different DMUs. The data of R&D, PC, SI, and EPS is collected from the third quarter financial statement of 2016 for these manufactures, which is different from the prediction of security companies for PI value source. This divergence of data sources brings different data attributes as the deterministic former ones and the non-deterministic latter one, which calls for the application of robust DEA models for their strength in dealing with uncertain data issues.

The raw data of six variables for 13 manufacturers is displayed in Table 1, from which we can see that the unit of R&D, PC, SI, PI, and RC is "Ten-thousand Yuan", while EPS is measured by "Yuan Per Share". What's more, some manufacturers haven't yet invested in R&D and some others are even in a deficit EPS situation, this higher variance sets a higher threshold to select a proper evaluation method. Fortunately, the new constructed SOCPR-DEA model has a conspicuous strength on translation invariance property, which makes it

could be well equipped to solve such issues with unprecise and inconsistent data. Taking BYD as an example, there are three different predicted PI values in 2016 as 11,736 thousands yuan from China Ping An Insurance (Group) Company Limited (PINGAN), 15,343.4 thousands yuan from Soochow Securities Company Limited (SCS), and 10,605.2 thousands yuan from UBS Securities Company Limited (UBS). To smooth the variance among three predictions, we adopt the method of arithmetic average to get the mean value of 12,561.5 thousands yuan and a 2020.5 thousands yuan variance.

4.1 The efficiency analysis for 13 NEV manufacturers

By using the SOCPR-DEA model to calculate the input–output ratio of different manufacturers, their efficiencies is shown in the last column of Table 1, from which we can see that five manufacturers as GWM, LIFAN, YUTONG, ZHONGTONG and ASIASTAR have the most efficient performances for their perfect efficiency values (the values equal to 1), three manufacturers as DFM, ANKAI, and BYD have a more efficient performances for their higher efficiency values (more than 0.4), while SAIC ranks as the last one for a 0.0386 efficiency value. It may be a surprising finding that SAIC, the largest automobile manufacturer of China has the lowest efficiency value because of the scale economies effect and its huge net profit. Nevertheless, if we focus on the NEV area of SAIC by comparing with its huge R&D investment and moderate sales revenue, our research finding is consistent with its practical performance. As the first automobile manufacturer in China which develops all the three electric vehicle technology roadmaps (the pure electric vehicle, the plug-in hybrid electric vehicle, and the fuel cell electric vehicle), SAIC has invested more than six billions yuan on R&D from 2009 to 2015,[2] however, this was returned by only about two thousands sales in 2014 and ten thousands sales in 2015.[3]

4.2 The sensitivity analysis among the efficiency and different variables

To excavate the relationship between the total efficiency and different variables, a sensitivity analysis was carried out through re-calculating the efficiencies of DMUs after deleting one variable at a time. The following Fig. 2 is constituted by 6 subgraphs, from left top to right bottom these 6 subgraphs are listed from subgraph 1 to subgraph 6. Subgraph 1 represent the default scenario including all the five variables, and subgraphs 2 to subgraph 6 represent the efficiency variance by eliminating R&D, PC, SI, EPS, and PI respectively. Figure 2 shows that the elimination of R&D, SI, EPS, and PI doesn't change the corresponding efficiency so much, while the efficiency is more sensitive with PC, and this is further confirmed by a rank-sum test displayed in Table 2. Since the null hypothesis is that the efficiency change caused by variable elimination is not significant, it is rejected only by PC variable with the 5% confidence through the *P* value judgement, while other variables have to accept the null hypothesis. Taking the variable of R&D as an example, the test shows that more R&D investment would not result in a higher efficiency. In fact, as Chinese NEV industry is still at an early stage, there is a long way to go before R&D outputs being transformed into technology capabilities or even marketing sales.

---

[2] Data source: http://www.ccstock.cn/gscy/gongsi/2016-11-03/A1478105064141.html.

[3] Data source: http://finance.qq.com/a/20160114/053577.htm.

Table 1 Raw data and performance efficiency of 13 large NEV manufactures in China

DMU	Main product	Input		Output				Efficiency
		R&D	PC	SI	EPS	PI		
SAIC	Passenger car and commercial vehicle	1210.43	45,879,100.00	52,385,500.00	2.09	N(7226.6, 88.33)		0.04
GWM	Sport utility vehicle	0.00	4,750,080.00	6,345,410.00	0.79	N(877.61, 26.8)		1.00
GAC	Passenger car	210,135.00	2,734,770.00	3,438,720.00	0.87	N(557.35, 102.99)		0.25
CHANGAN	Passenger car	104,097.00	4,439,730.00	5,360,700.00	1.66	N(759.28, 24.47)		0.19
BYD	Passenger car and commercial vehicle	250,071.00	5,745,410.00	7,279,780.00	1.38	N(1256.15, 202.05)		0.40
YUTONG	Bus	0.00	1,609,300.00	2,165,920.00	1.03	N(348.04, 12.75)		1.00
DFM	Commercial vehicle and passenger car	0.00	890,512.00	1,047,610.00	0.11	N(170.22, 10.64)		0.52
JAC	Commercial vehicle and passenger car	96,387.70	3,441,040.00	3,812,990.00	0.54	N(545.93, 27.03)		0.11
LIFAN	Family car	21,001.40	667,508.00	785,410.00	0.15	N(149.69, 16.27)		1.00
ANKAI	Bus	0.00	342,345.00	289,692.00	0.04	N(50.59, 0)		0.44
KING LONG	Bus	0.00	1,385,690.00	1,590,050.00	−0.17	N(284.13, 0)		0.07
ZHONGTONG	Bus	0.00	460,262.00	582,600.00	0.77	N(89.75, 5.15)		1.00
ASIASTAR	Coach	1.88	154,692.00	177,300.00	0.20	N(28.99, 0)		1.00
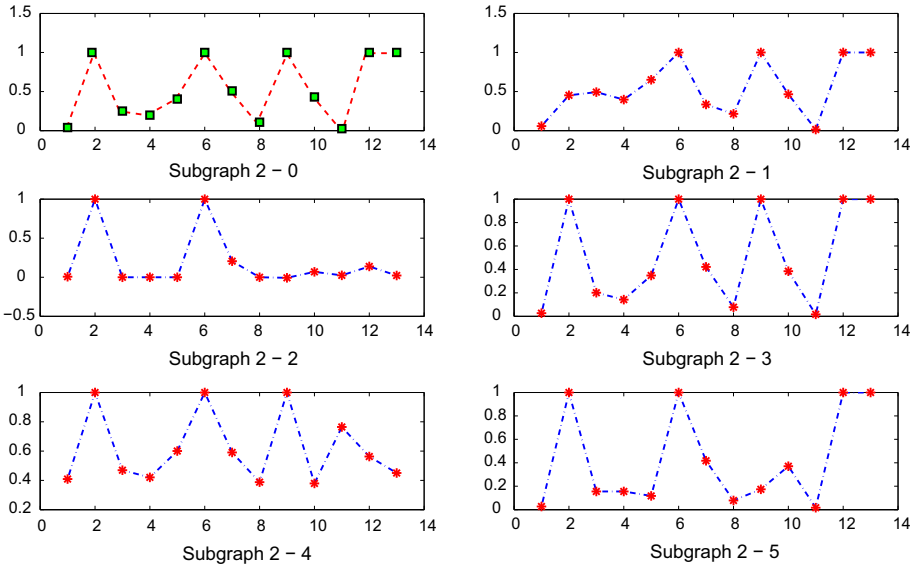
Fig. 2 Sensitivity analysis of variables to the efficiency

Table 2 Sensitivity analysis for each variable

Variable eliminated	R&D	PC	SI	EPS	PI
Sig.	0.7196	0.0041	0.6816	0.4119	0.2592

4.3 Practical contribution

The practical contributions of this paper are concluded below: Regarding the efficiencies of 13 NEV manufacturers, GWM (focusing on Sport Utility Vehicles), LIFAN (focusing on family cars), YUTONG (focusing on buses), ZHONGTONG (focusing on buses), and AISASTAR (focusing on coaches) have the largest efficiency value, while SAIC, BYD, and DFM with diverse products show lower efficiency performance. This phenomenon indicates that a focus strategy is more likely to enhance a firms efficiency especially at an emerging stage because of the absence of scale economies effect.

Regarding the sensitivity of different variables, PC reduction is the most effective way to improve a firm's efficiency. This can be realized through the following two measures:

– manufacturers could specialize in a small number of competitive products to upgrade production equipment and optimize manufacturing process, and expand its marketing channels to attract a large group of customers, so as to quickly form the scale economies effect;
– manufacturers could invest more on R&D to improve their technologies of both key components and NEVs for the lower sensitivity of R&D, so as to form their core competitiveness through the cumulative effect.

From the perspective of policy support, the comparison among GWM, LIFAN, YUTONG, ZHONGTONG, AISATAR, SAIC, BYD, and DFM ironically reflects that a prudential, focused and resource-based strategy is better for a higher firm efficiency due to the low R&D capability, inconsistent technical standards, and immature infrastructure facilities. Thus, we

suggest that the former comprehensive financial subsidy policy should be diverted to such critical areas as key technologies R&D, intellectual property protection, and charging facilities construction in the nearest future.

5 Conclusions and research prospects

5.1 Conclusions

To evaluate the economic performance for firms struggling in an emerging industry, this paper concerned more on data variety issues, proposed a second-order cone based robust data envelopment analysis (SOCPR-DEA) model, and carried out an empirical study by using 13 manufacturer cases from Chinese NEV industry. To elaborate, this paper has the following theoretical contributions.

Exploring a new robust DEA model: comparing with traditional DEA models, this paper developed a SOCPR-DEA model to evaluate the huge emerging uncertain issues in our real-life environment. Based on the M-DEA model, this paper firstly constructed a RM-DEA model framework through absorbing the robust optimization technique, and then transformed RM-DEA model into the final SOCPR-DEA model by integrating with the chance-constrained DEA model. The SOCPR-DEA model, which is computationally cheap, can effectively address imprecise and negative data problems and easily calculate efficiency evaluation issues by means of many mature software via inner point methods (e.g. CVX).

Taking advantage of the ellipsoidal uncertainty set: although the linear programming based classical robust models have a cheap computation virtue, it neglects the complex properties of uncertainty sets. To make up this defect, the SOCPR-DEA model proposed in this paper has completely utilized the advantages of ellipsoidal uncertainty set to well represent the uncertainties in our real-world. On the one hand, ellipsoidal uncertainty sets could form a relatively wide family including polytopes (bounded sets given by finitely linear inequalities), and this can be used to well approximate many cases of complicated convex sets. On the other hand, an ellipsoid is given parametrically by the moderate size data; hence it is convenient to represent "ellipsoidal uncertainty" as the input.

5.2 Future directions

There are some flaws in the paper such as the non-exhaustive consideration about the inner network of production processes for our focus on dealing with data variety issues; however, to combine the SOPER-DEA model with network DEA models deserves further investigates. Secondly, this research only focuses on Chinese NEV industry, other emerging industries such as hydraulic, photovoltaic, wind, ocean and solar energy could be evaluated more in the future, so as to make a comparison among different industries, and generalize more interesting implications for policy-makers reference. In addition, due to the imperfect statistical data and inconsistent statistical standards for different NEV manufacturers, we can only select six variables to carry out the efficiency evaluation, the validity deficiency could be mitigated in future studies by improving data collection and conducting more extensive fieldwork.

References

Aparicio, J., Mahlberg, B., Pastor, J. T., & Sahoo, B. K. (2014). Decomposing technical inefficiency using the principle of least action. *European Journal of Operational Research*, *239*(3), 776–785.

Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization* (Vol. 28). Princeton: Princeton University Press.

Ben-Tal, A., & Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, *23*(4), 769–805.

Ben-Tal, A., & Nemirovski, A. (1999). Robust solutions of uncertain linear programs. *Operations Research Letters*, *25*(1), 1–13.

Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations Research*, *52*(1), 35–53.

Boyd, S., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge: Cambridge University Press.

Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications*, *98*(2), 351–364.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444.

Chen, W., Gai, Y., & Gupta, P. (2018). Efficiency evaluation of fuzzy portfolio in different risk measures via DEA. *Annals of Operations Research*, *269*(1–2), 103–127.

Chen, K., & Zhu, J. (2019). Computational tractability of chance constrained data envelopment analysis. *European Journal of Operational Research*, *274*(3), 1037–1046.

Cooper, W. W., Thompson, R. G., & Thrall, R. M. (1996). Introduction: Extensions and new developments in DEA. *Annals of Operations Research*, *66*(1), 1–45.

El Ghaoui, L., & Lebret, H. (1997). Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, *18*(4), 1035–1064.

Emrouznejad, A., & Gl, Y. (2018). A survey and analysis of the first 40 years of scholarly literature in dea: 1978–2016. *Socio-Economic Planning Sciences*, *61*, 4–8.

Estellita Lins, M., Angulo-Meza, L., & Moreira da Silva, A. (2004). A multi-objective approach to determine alternative targets in data envelopment analysis. *Journal of the Operational Research Society*, *55*(10), 1090–1101.

Gharakhani, M., Kazemi, I., & Haji, H. (2011). A robust dea model for measuring the relative efficiency of Iranian high schools. *Management Science Letters*, *1*(3), 389–404.

Laguna, M. (1998). Applying robust optimization to capacity expansion of one location in telecommunications with demand uncertainty. *Management Science*, *44*(11–part–2), S101–S110.

Lee, A. H., Lin, C. Y., Kang, H. Y., & Lee, W. H. (2012). An integrated performance evaluation model for the photovoltaics industry. *Energies*, *5*(4), 1271–1291.

Leung, S., Wu, Y., & Lai, K. (2002). A robust optimization model for a cross-border logistics problem with fleet composition in an uncertain environment. *Mathematical and Computer Modelling*, *36*(11–13), 1221–1234.

Liu, W., Tao, M., & Zhao, X. (2018). Application of user research method based on user experience in new energy vehicle innovation. *Automobile Parts, 3*, 43–46. (**In Chinese**).

Liu, Y., & Kokko, A. (2013). Who does what in Chinas new energy vehicle industry? *Energy Policy*, *57*, 21–29.

Lu, C., Liu, H. C., Tao, J., Rong, K., & Hsieh, Y. C. (2017). A key stakeholder-based financial subsidy stimulation for Chinese EV industrialization: A system dynamics simulation. *Technological Forecasting and Social Change*, *118*, 1–14.

Lu, C., Rong, K., You, J., & Shi, Y. (2014). Business ecosystem and stakeholders' role transformation: Evidence from Chinese emerging electric vehicle industry. *Expert Systems with Applications*, *41*(10), 4579–4595.

Malcolm, S. A., & Zenios, S. A. (1994). Robust optimization for power systems capacity expansion under uncertainty. *Journal of the Operational Research Society*, *45*(9), 1040–1049.

Mulvey, J. M., Vanderbei, R. J., & Zenios, S. A. (1995). Robust optimization of large-scale systems. *Operations Research*, *43*(2), 264–281.

Omrani, H., Shafaat, K., & Emrouznejad, A. (2018). An integrated fuzzy clustering cooperative game data envelopment analysis model with application in hospital efficiency. *Expert Systems with Applications*, *114*, 615–628.

Sadjadi, S., & Omrani, H. (2008). Data envelopment analysis with uncertain data: An application for Iranian electricity distribution companies. *Energy Policy*, *36*(11), 4247–4254.

Shokouhi, A. H., Hatami-Marbini, A., Tavana, M., & Saati, S. (2010). A robust optimization approach for imprecise data envelopment analysis. *Computers & Industrial Engineering*, *59*(3), 387–397.

Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, *21*(5), 1154–1157.

Sueyoshi, T., Yuan, Y., & Goto, M. (2017). A literature study for dea applied to energy and environment. *Energy Economics*, *62*, 104–124.

Talluri, S., Narasimhan, R., & Nair, A. (2006). Vendor performance with supply risk: A chance-constrained dea approach. *International Journal of Production Economics*, *100*(2), 212–222.

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, *130*(3), 498–509.

Tone, K., & Tsutsui, M. (2010). Dynamic DEA: A slacks-based measure approach. *Omega*, *38*(3–4), 145–156.

Vassiadou, Z. C., & Zenios, S. A. (1996). Robust optimization models for managing callable bond portfolios. *European Journal of Operational Research*, *91*(2), 264–273.

Wen, M. (2015). *Uncertain data envelopment analysis*. Berlin: Springer.

Yan, H., & Wei, Q. (2011). Data envelopment analysis classification machine. *Information Sciences*, *181*(22), 5029–5041.

You, Y. Q., & Jie, T. (2016). A study of the operation efficiency and cost performance indices of power-supply companies in China based on a dynamic network slacks-based measure model. *Omega*, *60*, 85–97.

Yu, C. S., & Li, H. L. (2000). A robust optimization model for stochastic logistic problems. *International Journal of Production Economics*, *64*(1–3), 385–397.

Zhang, L., & Qin, Q. (2018). China's new energy vehicle policies: Evolution, comparison and recommendation. *Transportation Research Part A: Policy and Practice*, *110*, 57–72.

Zhou, Z., Lin, L., Xiao, H., Ma, C., & Wu, S. (2017). Stochastic network DEA models for two-stage systems under the centralized control organization mechanism. *Computers & Industrial Engineering*, *110*, 404–412.