International Journal of Innovative Computing, Information and Control
Volume **15**, Number **2**, April **2019**

ICIC International ©2019 ISSN 1349-4198
pp. 535–549

IMPROVED TWO-STAGE DEA MODEL: AN APPLICATION TO LOGISTICS EFFICIENCY EVALUATION ENTERPRISE IN XIAMEN, CHINA

Chuanzhong Yin[1], Wenhui Gao[1,*], Zhongheng Li[1], Zhongkai Wu[2] and Yuan Wang[1]

[1]College of Transport and Communications
Shanghai Maritime University
No. 1550, Haigang Ave., Shanghai 201306, P. R. China
czyin@shmtu.edu.cn; *Corresponding author: gwhsmwt@163.com
1048427185@qq.com; jerry_wang1996@yeah.net

[2]Operation and Development Department
China Railway Harbin Bureau Group Co., Ltd.
No. 51, West Dazhi Street, Harbin 150000, P. R. China
wzhk1972@126.com

Received June 2018; revised October 2018

Abstract. *With the rapid development of China's logistics industry, the efficiency of logistics enterprises has always been a significant concern of the Chinese government. Based on comprehensive studies on operation efficiency evaluation of logistics enterprises at home and abroad, we select the Data Envelopment Analysis (DEA) method to evaluate management efficiency in urban logistics enterprises. Traditional DEA model does not consider the internal operation process, and its decision-making units (expressed as DMU or DMUs in below) are regarded as a "black box", and traditional two-stage DEA method opened the "black box", but only considered the original inputs and final outputs of DMUs, without addressing the structural problem of intermediate products. We improved two-stage DEA model, using the shared inputs of two-stages, direct outputs of intermediate products and additional inputs at the second stage to improve the index structure, so as to avoid the shortcomings that traditional DEA model may be overscored and traditional two-stage DEA model may be underscored. We take Xiamen city of China as an example to carry out an empirical study, because the relative quantity of indicators and enterprises in the data we used conflicts with "freedom degree" requirement in DEA method, prior to evaluation, we used correlation analysis and improved principal component analysis to eliminate the indicators that reflect information overlap and have little impact on the evaluation result, and the result shows that the index screening method is feasible and reliable.*

Keywords: Logistics enterprises, Efficiency evaluation, Improved two-stage DEA model, Correlation analysis, Improved principal component analysis

1. **Introduction.** As the third profit source, logistics is gradually emphasized after natural resources consumption is reduced and social labor efficiency is enhanced. Nowadays mobility of social resources is so extensive and active that logistics industry plays a significant role. Yet with development of economy, logistics industry is no longer a new industry, the development of which is intermingled. How to promote the healthy and steady development of logistics industry is a key issue confronting each country. In China, the development plan of logistics industry has been introduced on the national and local level, such as *Logistics Industry Restructuring & Revitalization Plan*, *Logistics Plan in*

DOI: 10.24507/ijicic.15.02.535

the 13th Five-Year (Xiamen is one of the node cities), in order to promote the sustainable development of logistics industry. It is significant to evaluate the efficiency of logistics industry scientifically, understand its development status and assess its tendency accurately for supporting scientific decisions of logistics development missions and guide the trend of logistics industry.

In this study, Xiamen, which is one of Chinese coastal cities with more active logistics development, is taken as an example to study the operation efficiency of logistics enterprises in the whole city. Through scientific index system and evaluation model, we can know the logistics efficiency from different industries, provide decision support for government departments to propose targeted logistics policies and plans, and provide a valuable reference for the development of enterprises.

The research adopts a two-stage DEA model for efficiency evaluation of logistics enterprises. The total efficiency of DMUs can be assessed by the model and decomposed into an efficiency value for each stage. Specifically, this study first expounds the evolution process of the DEA model and analyzes its shortcomings, and then puts forward the general idea of this study. Secondly, it constructs the improved two-stage DEA model and the overall efficiency of the DMU under the two-stage DEA system considering the internal subprocess; furthermore, in order to improve the credibility of the evaluation system, the three methods, that is, the preliminary screening of indicators, the index screening based on correlation analysis and the improved principal component analysis, are selected. Finally, we use the data of Xiamen logistics enterprises to carry out an example analysis. Therefore, the model can provide internal operation of DMUs with more accurate information. Meanwhile, in the field of theoretical development of DEA method, establishing a systematic operation model considering input and intermediate products is a heated topic in current researches.

The innovation of this paper is to classify the logistics enterprises scientifically and establish the efficiency evaluation index system of logistics enterprises. Through the correlation analysis method and the improved principal component analysis method, the excessive indexes have been sifted down and allocated reasonably. The improved two-stage DEA method can effectively overcome the shortcomings of the traditional DEA method. Based on the empirical analysis of the efficiency of Xiamen logistics enterprises in China, this study obtains the efficiency of the five types of logistics enterprises and has a strong practical application value for the enterprise operation decision and the government's development planning of the logistics industry.

2. **Literature Review.** Since DEA method proposed by American renowned operational researchers Charnes et al. [1], the model has been widely used in various industries and departments and have advantages in processing multiple inputs and multiple outputs. Meanwhile, the model is input-oriented and assumes Constant Returns to Scale (CRS) and the model is also called CCR (A. Charnes & W. W. Cooper & E. Rhodes) model or C^2R model. In traditional DEA model, internal operations during the process are not considered. As a "Black Box", original inputs and final outputs of DMUs are considered by a single process. However, accurate information of ineffective resources of DMUs cannot be provided if the internal structure is neglected. To solve this problem, researchers have conducted many researches on decomposing total efficiency. For instance, Banker et al. [2] decomposed efficiencies into scale efficiencies and technical efficiencies; subsequently, Byrnes and Grosskopf [3] identified congestion effect from technical efficiencies; Kao [4] obtained total efficiency through calculating weighted average of single output value. The methods decompose efficiency of DMUs into various parts to gain ineffective information about DMUs; Choi and Ahn [5] solved the problem that the traditional DEA model

does not provide the grouping information on the efficient units or inefficient units, and proposed a new approach based on DEA and clustering model. The above literature depends on the structure of traditional DEA model, not actually opening the "Black Box". Färe and Whittaker [6] proposed multistage DEA model, decomposing a production process into several subprocesses. An output from one subprocess is used as an input to another. Subsequently, Färe and Grosskopf [7] also proposed the concept of network DEA, essence of which is to open the "black box" to examine the efficiency of each component of production process and its impact on overall efficiency of the system. They also pointed out that the multistage DEA model is a special case of the network DEA model. The above literature improves the deficiency of the model, opens the internal structure, but is still based on the traditional DEA model, and the disadvantage is that it mainly emphasizes the internal structure of DMUs and interdependence of different subprocesses.

Some two-stage models about internal structure of DEA were established in recent years, such as Castelli et al. [8], Cook et al. [9] and Kao [10]. Among the research on internal structure, an ordinary internal structure is considered as traditional two-stage network process, where outputs from the first stage become inputs to the second stage. Traditional two-stage model mainly emphasizes the internal structure of DMUs or interdependence of different subprocesses. If input or intermediate products have a complicated structure, effectiveness of the model will be confined. According to this problem, some scholars allocated inputs to different stages to improve the model limitation and proposed to add some inputs in the second stage. For example, Yu and Fan [11], Zha [12], and Chen et al. [13] proposed a two-stage DEA model of shared input for the first and second stage; Liang et al. [14] and Chen et al. [15] proposed a two-stage model to add inputs at the second stage. However, although these studies have considered structural problems in the construction of DEA model, allocations of intermediate products are neglected in the effective evaluation and decomposition process.

Lozano et al. [16] and Maghbouli et al. [17] established a two-stage DEA model, considering both intermediate products and bad output produced by the first stage. Yu and Shi [18] established a model, adding inputs to the second stage and taking intermediate product as the final output, implementing the free distribution of intermediate products and analyzing the problem of maximizing system efficiency by cooperative game. Although these studies have made some improvements on the basis of traditional two-stage model, for three conditions, i.e., shared inputs of two-stages, newly added inputs to the second stage and direct outputs from intermediate products, only one or two conditions are considered.

Referring to the study of Ma [19], this research improves intermediate products of DMUs in the model of efficiency evaluation. By adopting a two-stage DEA model, one part of intermediate products are inputted to the second stage while the other part as final outputs. Compared with the previous literature, the paper not only improves the limitation of the model but also emphasizes the distribution of intermediate products in the effective evaluation and decomposition process. Meanwhile, the model considers the input structure including original inputs to the first stage, shared inputs of two-stages and newly added inputs to the second stage. Thus, results of evaluation are more scientific and practical.

3. **Improved Two-Stage DEA Model.**

3.1. **Structure of improved two-stage DEA model.** Consider a series of decision-making units DMUs. Suppose each DMU_j $(j = 1, 2, \ldots, n)$ joins in the two-stage network process, shown in Figure 1.

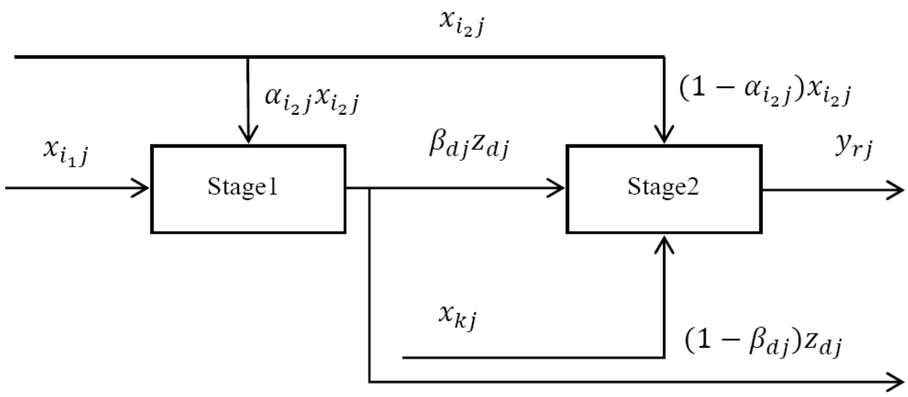


Figure 1. Structure of improved two-stage DEA model

Suppose each DMU_j has two types of inputs: m original inputs, x_{ij} $(i = 1, 2, \ldots, m)$, and q additional inputs, x_{kj} $(k = 1, 2, \ldots, q)$. Only one part of original inputs x_{i_1j} $(i_1 \in I_1)$ are the inputs to the first stage while the remaining parts x_{i_2j} $(i_2 \in I_2)$ are shared inputs of two-stages, $I_1 \cup I_2 = \{1, 2, \ldots, m\}$ and $I_1 \cap I_2 = \varnothing$. As the constraints in Cook and Hababou [20], the portion α_{i_2j} $(i_2 \in I_2,\ j = 1, \ldots, n)$ should be within certain intervals, noted $L^1_{i_2j} \le \alpha_{i_2j} \le L^2_{i_2j}$.

Suppose each DMU_j has D outputs z_{dj} $(d = 1, 2, \ldots, D)$ from the first stage and s outputs y_{rj} $(r = 1, 2, \ldots, s)$ from the second stage. Part of z_{dj} are intermediate outputs or inputs to the second stage while the remaining of z_{dj} are final outputs. Intermediate part is $\beta_{dj}z_{dj}$ and the part of the direct output, no longer as input of the second stage, is $(1 - \beta_{dj})z_{dj}$, $0 < \beta_{dj} \le 1$. Similarly, β_{dj} is within the interval of $H^1_{dj} \le \beta_{dj} \le H^2_{dj}$.

3.2. **Improved two-stage DEA model.** Consider the two-stage process shown in Figure 1, $0 < \alpha_{i_2j} < 1$, $0 < \beta_{dj} < 1$. To be more exactly, under this situation, two decisions have to be made for each DMU: resource allocation and redistribution of intermediate products. During the two-stage process of shared inputs, free intermediate products and the new input at the second stage, relative efficiency of DMU_0 at the first stage is θ_0^{SF1} and that at the second stage is θ_0^{SF2}, which can be obtained by the following two models:

$$
\begin{aligned}
\theta_0^{SF1} &= \max \frac{\sum_{d=1}^{D} u_d^1 z_{d0}}{\sum_{i_1\in I_1} \nu_{i_1} x_{i_10} + \sum_{i_2 \in I_2} \nu_{i_2}^1 \alpha_{i_20} x_{i_20}} \\
\text{s.t.} \quad & \frac{\sum_{d=1}^{D} u_d^1 z_{dj}}{\sum_{i_1\in I_1} \nu_{i_1} x_{i_1j} + \sum_{i_2 \in I_2} \nu_{i_2}^1 \alpha_{i_2j} x_{i_2j}} \le 1, \quad j = 1, \ldots, n \\
& L^1_{i_2j} \le \alpha_{i_2j} \le L^2_{i_2j}, \quad i_2 \in I_2, \quad j = 1, \ldots, n \\
& u_d^1, \nu_{i_1}, \nu_{i_2}^1 \ge \varepsilon, \quad d = 1, \ldots, D, \quad i_1 \in I_1,\ i_2 \in I_2
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
\theta_0^{SF2} &= \max \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i_2\in I_2} \nu_{i_2}^2 (1-\alpha_{i_20}) x_{i_20} + \sum_{k=1}^{q} \nu_k x_{k0} + \sum_{d=1}^{D} u_d^2 \beta_{d0} z_{d0}} \\
\text{s.t.} \quad & \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i_2\in I_2} \nu_{i_2}^2 (1-\alpha_{i_2j}) x_{i_2j} + \sum_{k=1}^{q} \nu_k x_{kj} + \sum_{d=1}^{D} u_d^2 \beta_{dj} z_{dj}} \le 1, \quad j = 1, \ldots, n \\
& L^1_{i_2j} \le \alpha_{i_2j} \le L^2_{i_2j},\ i_2 \in I_2,\ j = 1, \ldots, n \\
& H^1_{dj} \le \beta_{dj} \le H^2_{dj},\ d = 1, \ldots, D,\ j = 1, \ldots, n \\
& u_r, u_d^2, \nu_{i_2}^2, \nu_k \ge \varepsilon,\ r = 1, \ldots, s,\ d = 1, \ldots, D,\ k = 1, \ldots, q,\ i_1 \in I_1,\ i_2 \in I_2
\end{aligned} \tag{2}
$$

where ν_{i_1} are the weight associated with initial inputs i_1; $\nu_{i_2}^1$, $\nu_{i_2}^2$ are respectively the weights associated with shared inputs i_2 in the first stage and second stage; u_d^1, u_d^2 are the weights associated with outputs d from the first stage and as the inputs in the second stage; ν_k are the weight associated with additional inputs k; and u_r are the weight associated with final outputs r.

Even if there are only parts of the outputs of the first stage which become the inputs of the second stage, the value accorded these intermediate measures should reasonably be assumed as identical in the two stages. Moreover, $\forall i_2 \in I_2$, x_{i_2} is the same inputs for the whole process, and the weights of these inputs can be supposed to be equal in each stage. We can therefore assume that $u_d^1 = u_d^2 = u_d$ $(d = 1, \ldots, D)$ and $\nu_{i_2}^1 = \nu_{i_2}^2 = \nu_{i_2}$ $(i_2 \in I_2)$ in models (1) and (2). Calculation formula of total efficiency of two-stages can be defined as $\theta_0^{SF} = w_1\theta_0^{SF1} + w_2\theta_0^{SF2}$, in which:

$$w_1 = \frac{\sum_{i_1\in I_1} \nu_{i_1}x_{i_10} + \sum_{i_2\in I_2} \nu_{i_2}\alpha_{i_20}x_{i_20}}{\sum_{i_1\in I_1} \nu_{i_1}x_{i_10} + \sum_{i_2\in I_2} \nu_{i_2}x_{i_20} + \sum_{k=1}^{q} \nu_k x_{k0} + \sum_{d=1}^{D} u_d\beta_{d0}z_{d0}} \tag{3}$$

$$w_2 = \frac{\sum_{i_2\in I_2} \nu_{i_2}(1-\alpha_{i_20})x_{i_20} + \sum_{k=1}^{q} \nu_k x_{k0} + \sum_{d=1}^{D} u_d\beta_{d0}z_{d0}}{\sum_{i_1\in I_1} \nu_{i_1}x_{i_10} + \sum_{i_2\in I_2} \nu_{i_2}x_{i_20} + \sum_{k=1}^{q} \nu_k x_{k0} + \sum_{d=1}^{D} u_d\beta_{d0}z_{d0}} \tag{4}$$

The fractional programming model of total efficiency of DMU_0 is:

$$\begin{aligned}
&\theta_0^{SF} = \frac{\sum_{d=1}^{D} u_d z_{d0} + \sum_{r=1}^{s} u_r y_{r0}}{\sum_{i_1\in I_1} \nu_{i_1}x_{i_10} + \sum_{i_2\in I_2} \nu_{i_2}x_{i_20} + \sum_{k=1}^{q} \nu_k x_{k0} + \sum_{d=1}^{D} u_d\beta_{d0}z_{d0}}\\
\text{s.t.}\ &\frac{\sum_{d=1}^{D} u_d z_{dj}}{\sum_{i_1\in I_1} \nu_{i_1}x_{i_1j} + \sum_{i_2\in I_2} \nu_{i_2}\alpha_{i_2j}x_{i_2j}} \le 1, \quad j = 1, \ldots, n\\
&\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i_2\in I_2} \nu_{i_2}(1-\alpha_{i_2j})x_{i_2j} + \sum_{k=1}^{q} \nu_k x_{kj} + \sum_{d=1}^{D} u_d\beta_{dj}z_{dj}} \le 1, \quad j = 1, \ldots, n\\
&L_{i_2j}^1 \le \alpha_{i_2j} \le L_{i_2j}^2,\ i_2 \in I_2,\ j = 1, \ldots, n\\
&H_{dj}^1 \le \beta_{dj} \le H_{dj}^2,\ d = 1, \ldots, D,\ j = 1, \ldots, n\\
&u_r, u_d, \nu_{i_1}, \nu_{i_2}, \nu_k \ge \varepsilon,\ r = 1, \ldots, s,\ d = 1, \ldots, D,\ k = 1, \ldots, q,\ i_1 \in I_1,\ i_2 \in I_2
\end{aligned} \tag{5}$$

Let $t = \frac{\sum_{d=1}^{D} u_d z_{d0} + \sum_{r=1}^{s} u_r y_{r0}}{\sum_{i_1\in I_1} \nu_{i_1}x_{i_10} + \sum_{i_2\in I_2} \nu_{i_2}x_{i_20} + \sum_{k=1}^{q} \nu_k x_{k0} + \sum_{d=1}^{D} u_d\beta_{d0}z_{d0}}$, and define $\mu_d = tu_d$, $\mu_r = tu_r$, $\upsilon_{i_1} = t\nu_{i_1}$, $\upsilon_{i_2} = t\nu_{i_2}$, $\upsilon_k = t\nu_k$, $\omega_{i_2j} = \upsilon_{i_2}\alpha_{i_2j}$ and $\eta_{dj} = \mu_d\beta_{dj}$. Model (5) is equal to the following linear programming:

$$\begin{aligned}
&\theta_0^{SF} = \max \sum_{d=1}^{D} \mu_d z_{d0} + \sum_{r=1}^{s} \mu_r y_{r0}\\
\text{s.t.}\ &\sum_{d=1}^{D} \mu_d z_{dj} - \sum_{i_1\in I_1} \upsilon_{i_1}x_{i_1j} - \sum_{i_2\in I_2} \omega_{i_2j}x_{i_2j} \le 0, \quad j = 1, \ldots, n\\
&\sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i_2\in I_2} \upsilon_{i_2}x_{i_2j} + \sum_{i_2\in I_2} \omega_{i_2j}x_{i_2j} - \sum_{k=1}^{q} \upsilon_k x_{kj} - \sum_{d=1}^{D} \eta_{dj}z_{dj} \le 0,\ j = 1, \ldots, n\\
&\sum_{i_1\in I_1} \upsilon_{i_1}x_{i_10} + \sum_{i_2\in I_2} \upsilon_{i_2}x_{i_20} + \sum_{k=1}^{q} \upsilon_k x_{k0} + \sum_{d=1}^{D} \eta_{d0}z_{d0} = 1\\
&\mu_d, \mu_r, \eta_{dj}, \upsilon_{i_1}, \upsilon_{i_2}, \upsilon_k, \omega_{i_2j} \ge \varepsilon, \quad d = 1, \ldots, D,\ r = 1, \ldots, s,\\
&j = 1, \ldots, n,\ i_1 \in I_1,\ i_2 \in I_2,\ k = 1, \ldots, q
\end{aligned} \tag{6}$$

Model (6) provides the model of total efficiency of DMU under the two-stage DEA system with internal subprocesses shown by Figure 1.

3.3. **Efficiency decomposition of improved two-stage DEA model.** Model (6) considers both structures of inputs and intermediate products. Decomposition of efficiency can be obtained by the following model. When the first stage is endowed with priority, its efficiency can be obtained by the following formula:

$$
\begin{aligned}
&\theta_0^{SF1*} = \max \sum_{d=1}^{D} \mu_d z_{d0} \\
\text{s.t.}\ &\sum_{d=1}^{D} \mu_d z_{dj} - \sum_{i_1 \in I_1} \upsilon_{i_1} x_{i_1 j} - \sum_{i_2 \in I_2} \omega_{i_2 j} x_{i_2 j} \le 0, \quad j = 1, \ldots, n \\
&\sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i_2 \in I_2} \upsilon_{i_2} x_{i_2 j} + \sum_{i_2 \in I_2} \omega_{i_2 j} x_{i_2 j} - \sum_{k=1}^{q} \upsilon_k x_{kj} - \sum_{d=1}^{D} \eta_{dj} z_{dj} \le 0,\ j = 1, \ldots, n \\
&\sum_{d=1}^{D} \mu_d z_{d0} + \sum_{r=1}^{s} \mu_r y_{r0} - \theta_0^{SF*} \left(\sum_{i_1 \in I_1} \upsilon_{i_1} x_{i_1 0} - \sum_{i_2 \in I_2} \upsilon_{i_2} x_{i_2 0} \right. \\
&\left. + \sum_{k=1}^{q} \upsilon_k x_{k0} - \sum_{d=1}^{D} \eta_{d0} z_{d0} \right) = 0 \\
&\sum_{i_1 \in I_1} \upsilon_{i_1} x_{i_1 0} + \sum_{i_2 \in I_2} \omega_{i_2 0} x_{i_2 0} = 1 \\
&\mu_d, \mu_r, \eta_{dj}, \upsilon_{i_1}, \upsilon_{i_2}, \upsilon_k, \omega_{i_2 j} \ge \varepsilon, \quad d = 1, \ldots, D,\ r = 1, \ldots, s, \\
&j = 1, \ldots, n,\ i_1 \in I_1,\ i_2 \in I_2,\ k = 1, \ldots, q
\end{aligned}
\tag{7}
$$

And efficiency value of the second stage can be obtained by $\theta_0^{SF2} = \frac{\theta_0^{SF*} - w_1^* \theta_0^{SF1*}}{w_2^*}$. When the second stage is endowed with priority, its efficiency value can be obtained by the following:

$$
\begin{aligned}
&\theta_0^{SF2*} = \max \sum_{r=1}^{s} \mu_r y_{r0} \\
\text{s.t.}\ &\sum_{d=1}^{D} \mu_d z_{dj} - \sum_{i_1 \in I_1} \upsilon_{i_1} x_{i_1 j} - \sum_{i_2 \in I_2} \omega_{i_2 j} x_{i_2 j} \le 0, \quad j = 1, \ldots, n \\
&\sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i_2 \in I_2} \upsilon_{i_2} x_{i_2 j} + \sum_{i_2 \in I_2} \omega_{i_2 j} x_{i_2 j} - \sum_{k=1}^{q} \upsilon_k x_{kj} - \sum_{d=1}^{D} \eta_{dj} z_{dj} \le 0,\ j = 1, \ldots, n \\
&\sum_{d=1}^{D} \mu_d z_{d0} + \sum_{r=1}^{s} \mu_r y_{r0} - \theta_0^{SF*} \left(\sum_{i_1 \in I_1} \upsilon_{i_1} x_{i_1 0} - \sum_{i_2 \in I_2} \upsilon_{i_2} x_{i_2 0} \right. \\
&\left. + \sum_{k=1}^{q} \upsilon_k x_{k0} - \sum_{d=1}^{D} \eta_{d0} z_{d0} \right) = 0 \\
&\sum_{i_2 \in I_2} \upsilon_{i_2} x_{i_2 0} - \sum_{i_2 \in I_2} \omega_{i_2 0} x_{i_2 0} + \sum_{k=1}^{q} \upsilon_k x_{k0} + \sum_{d=1}^{D} \eta_{d0} z_{d0} = 1 \\
&\mu_d, \mu_r, \eta_{dj}, \upsilon_{i_1}, \upsilon_{i_2}, \upsilon_k, \omega_{i_2 j} \ge \varepsilon,\ d = 1, \ldots, D,\ r = 1, \ldots, s,
\end{aligned}
$$

$$j = 1, \dots, n,\ i_1 \in I_1,\ i_2 \in I_2,\ k = 1, \dots, q \tag{8}$$

And efficiency value of the first stage can be obtained by $\theta_0^{SF1} = \frac{\theta_0^{SF*} - w_2^* \theta_0^{SF2*}}{w_1^*}$.

Finally, unique efficiency decomposition can be obtained if $\theta_0^{SF1*} = \theta_0^{SF1}$ or $\theta_0^{SF2*} = \theta_0^{SF2}$.

4. **Evaluation Index of Logistics Enterprise.** This paper applies improved two-stage DEA method in study of operational efficiencies of logistics enterprises. Due to the discrepancy among different enterprises, there are more indicators to fill in, which ensures that data collection is more comprehensive to avoid missing information. However, selection of numerous indexes may lead to long-winded information or high relevance between indexes; as a consequence, unnecessary work may be added in the process of efficiency evaluation. Moreover, DEA evaluation model that we used has the requirement for "freedom degree", which means enough DMUs are needed. The relationship among the number of input indexes (denoted as M), the number of output indexes (denoted as N) and the number of DMUs (denoted as K) is $2(M + N) \leq K$; otherwise, confidence degree of evaluation results will drop. Therefore, when applying DEA method, the number of input or output variables should be appropriate. We abridged the stronger relevant indicators and eliminated the indicators that had less impact on the results of evaluation, which will reduce complexity of the research problem to ensure the effectiveness of evaluation work. Here, before the evaluation, we first select the indicators, and specific process is shown in Figure 2.

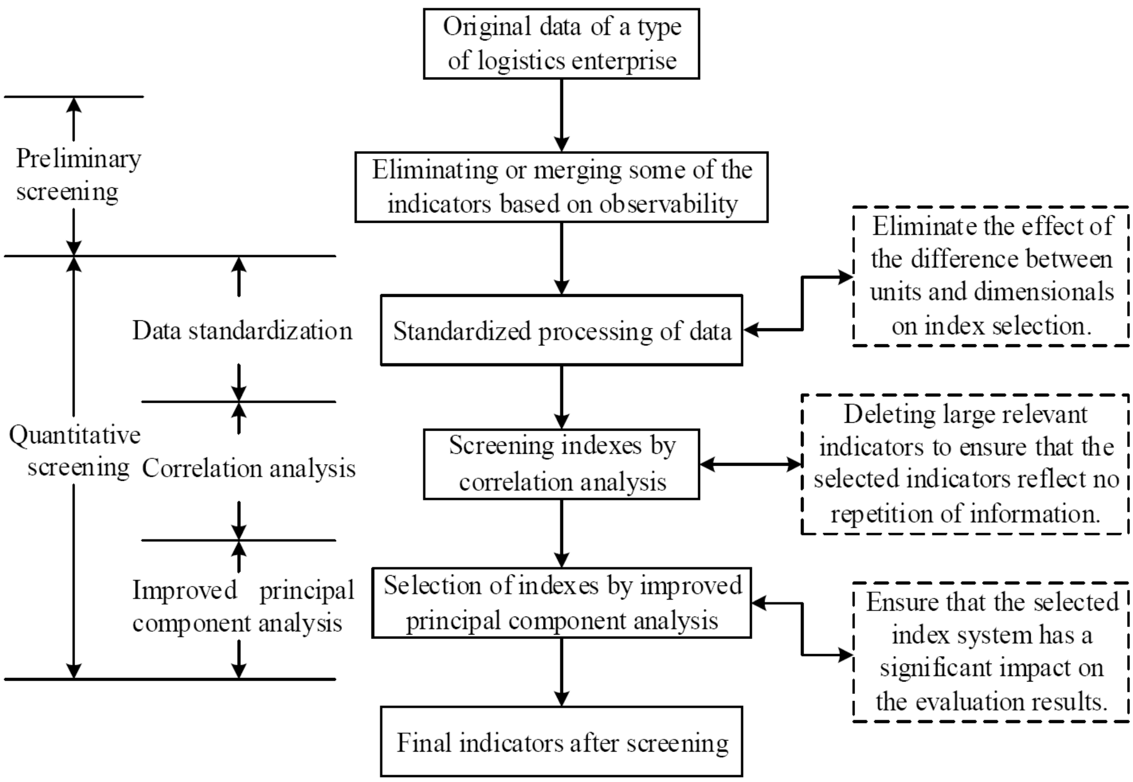


FIGURE 2. Flow chart of index screening

4.1. **Preliminary screening and index standardization.**

1) Preliminary screening of the index

Screen the index preliminarily based on observability, consider that number of indexes cannot exceed that of enterprises, delete or integrate indexes whose meanings overlap or include each other, in order to reduce dimensions of indexes under the condition of ensuring complete information to the greatest extent.

2) Standardization of index

The purpose of standardization is to remove dimensions, unify the standard of index and eliminate the impact on evaluation results. We standardize indexes to the interval of $[0, 1]$, which is positive standardization.

The following is the equation of positive standardization:

$$p_{ij} = \frac{\nu_{ij} - \min\limits_{1 \le j \le n}(\nu_{ij})}{\max\limits_{1 \le j \le n}(\nu_{ij}) - \min\limits_{1 \le j \le n}(\nu_{ij})} \tag{9}$$

In the equation, p_{ij} – standardization value of the i^{th} index and j^{th} evaluation object, ν_{ij} – observation value of the i^{th} index and j^{th} evaluation object, n – quantity of evaluation objects.

4.2. **Index screening based on correlation analysis.** Correlation coefficient reflects the relevance between two indexes. High correlation coefficient means redundancy of the index. That is to say, it is possible to reduce the highly relevant indicators on the premise of keeping roughly the same information content. Through correlation analysis, index system becomes more concise and information is ensured as complete as possible. The method of screening index based on correlation analysis is as follows.

1) Calculate correlation coefficient between each index

Calculation formula of correlation coefficient among different indexes is as follows:

$$r_{ij} = \frac{\sum_{k=1}^{n}(z_{ik} - \overline{z_i})(z_{jk} - \overline{z_j})}{\sqrt{\sum_{k=1}^{n}(z_{ik} - \overline{z_i})^2 (z_{jk} - \overline{z_j})^2}} \tag{10}$$

In the equation, r_{ij} is correlation coefficient between the i^{th} and j^{th} indexes; n is the evaluation object; z_{ik} is value of the i^{th} index and k^{th} evaluation object; z_{jk} is value of the j^{th} index and k^{th} evaluation object; $\overline{z_i}$ and $\overline{z_j}$ are respectively averages of the i^{th} and j^{th} indexes.

2) Define a critical value M $(0 < M < 1)$. When $|r_{ij}| > M$, delete an index; otherwise, keep both.

4.3. **Index screening based on improved principal component analysis.** Principal component analysis is the most commonly used in dimension reduction of index. However, the indexes extracted by principal component analysis are principal component variables. When the principal component variable is applied to efficiency evaluation of DEA method, effective frontier can only improve the principal component variable of non-effective unit involved in the operation. As a consequence, original inputs and outputs cannot be improved directly, leading to difficulties for managers finding problems. Thus, for this reason, we propose the improved principal component analysis based on principal component analysis.

4.3.1. *Principal component analysis.* Suppose n DMUs and each unit has m indexes. $X = (x)^T_{m \times n}$ is the data set. Specific steps of principal component analysis calculation are as follows.

Step 1: Standardize the sample data X and get standard data matrix, Y $= {y_{ij}}^T_{m\times n} = (Y_1, Y_2, \ldots, Y_m)^T$. Standardization equation is:

$$y_{ij} = \frac{x_{ij} - \overline{x_i}}{s_i} \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n) \tag{11}$$

$\overline{x_i}$ and S_i are respectively average and deviation of the sample data.

Step 2: Calculate correlation matrix R $= (r_{ij})_{m\times m}$ corresponding to standard matrix Y. Correlation coefficient is obtained by the following equation:

$$r_{ij} = \frac{1}{n-1}\sum_{t=1}^{n} y_{ti}y_{tj} \tag{12}$$

Step 3: Calculate eigenvalue of correlation matrix R, $\lambda_1, \lambda_2, \ldots, \lambda_m$ ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$), and corresponding feature vector, $l_1, l_2, \ldots, l_m$. Eigenvalue can be calculated by the eigenvalue equation, $|R - \lambda I| = 0$, $l_i = (l_{1i}, l_{2i}, \ldots, l_{mi})^T$, $j = 1, 2, \ldots, m$.

Step 4: Extract principal component according to cumulative contribution. Contribution rate of each component is $b_i = \lambda_i \left(\sum_{t=1}^{m} \lambda_t\right)^{-1}$ $(i = 1, 2, \ldots, m)$. According to the cumulative contribution rate principle:

$$\left(\sum_{t=1}^{k} \lambda_t\right)\left(\sum_{t=1}^{n} \lambda_t\right)^{-1} \geq 85\% \tag{13}$$

Select k principal components $Z_i = \sum_{t=1}^{m} l_{ti}Y_t$ $(i = 1, 2, \ldots, k)$ as the new input variables.

The final i^{th} principal component is linear combination of standard matrix, coefficient of which is the component of feature vector l_i corresponding to eigenvalue λ_i.

4.3.2. *Improved principal component analysis.* Similarly, suppose n DMUs and each unit has m inputs and s outputs. Dataset of input variables is X $= (x)^T_{m\times n}$ and that of output variables is Y $= (y_{rj})^T_{s\times j}$, Z $= (Z_1, Z_2, \ldots, Z_m)^T$ indicates m principal components. $(a_1, a_2, \ldots, a_m)^T$ indicates relative contribution value of m input variables. The modified principal component analysis can be combined with DEA method to evaluate the efficiency of decision unit more effectively. The followings are specific calculation steps:

Step 1: Calculate standardized matrix X* of input data X and corresponding relative matrix R. The methods are the same as above;

Step 2: Solve eigenvalue $\lambda_1, \lambda_2, \ldots, \lambda_m$ ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$) of relative matrix R by eigenvalue equation $|R - \lambda I| = 0$. Then solve corresponding feature vector $l_i = (l_1, l_2, \ldots, l_m)^T$ by homogeneous linear equation $(R - \lambda I)L = 0$ and calculate contribution rate $b_i = \lambda_i \left(\sum_{t=1}^{m} \lambda_t\right)^{-1}$, $(i = 1, 2, \ldots, m)$;

Step 3: Transform equation $Z = L^T X^*$ ($L = (l_1, l_2, \ldots, l_m)^T$) into $X^* = LZ$, then $(a_1, a_2, \ldots, a_m)^T = |L * (b_1, b_2, \ldots, b_m)^T|$.

Absolute value means the absolute value of m elements of column vector.

Step 4: Rank elements of vector $(a_1, a_2, \ldots, a_m)^T$ from big to small and obtain vector $(a_1^*, a_2^*, \ldots, a_m^*)^T$ ($a_1^* \geq a_2^* \geq \cdots \geq a_m^*$);

Step 5: According to cumulative contribution criteria, that is $\frac{\sum_{i=1}^{k} a_i^*}{\sum_{i=1}^{m} a_i^*} \geq 85\%$, extract k variables to complete the selection of m input variables;

Step 6: Selection of output variables can be completed by the same method.

As can be seen, from transformation of Step 3, several key variables can be selected from original indexes, and instead of a linear combination of original index, a new principal component is produced. Thus, original index can be directly involved in the operation of the DEA model.

5. **Case Study.** This paper takes the logistics enterprises of Xiamen as research objects. Data are collected from questionnaires. The survey collects 12 express enterprises, 15 warehousing enterprises, 21 third-party logistics (3PL) enterprises, 41 transportation enterprises and 26 freight forwarding enterprises, from December 2016 to February 2017. Specific evaluation process is indicated by the example of express enterprises.

5.1. Index processing.

5.1.1. *Index after preliminary screening.* After preliminary screening, indexes of express enterprises are shown in Table 1.

TABLE 1. Index of express-delivery enterprises after preliminary screening

Index	ID.1	ID.2	ID.3	ID.4	ID.5	ID.6
Meaning	Total Staff Number (Person)	Staff Number with Bachelor's Degree and Above (Person)	Own Vehicles (Unit)	Enterprise Management Software (Piece)	City Dot (Unit)	Area of Warehouse (Square Meter)
Index	ID.7	ID.8	ID.9	ID.10	ID.11	ID.12
Meaning	Total Business Revenue (10,000 Yuan)	Monthly Processing Order Quantity (Unit)	Quantity of Picking (Piece)	Delivery of Cargoes (Piece)	Monthly Transported Cargoes (Piece)	Monthly Transported Cargoes (Ton)

5.1.2. *Screening index based on correlation analysis.*

1) Calculate correlation coefficient.
2) Classify indexes of each category of enterprises into two types: input and output. Calculate correlation coefficient between indexes from the collection of input and output indexes. Take the express delivery enterprise as example. Correlation coefficient matrix can be obtained by Equation (10) with indexes of express delivery enterprise after preliminary screening. Result of calculation is indicated in Table 2.

TABLE 2. Correlation coefficient matrix of express-delivery enterprise

Correlation Coefficient Matrix of Input Indexes	ID.1	ID.2	ID.3	ID.4	ID.5	ID.6
ID.1	1	0.9352	0.6713	0.452	0.4397	0.799
ID.2	0.9352	1	0.5468	0.4292	0.5074	0.8278
ID.3	0.6713	0.5468	1	0.3367	0.2912	0.4395
ID.4	0.452	0.4292	0.3367	1	0.1709	0.084
ID.5	0.4397	0.5074	0.2912	0.1709	1	0.1584
ID.6	0.799	0.8278	0.4395	0.084	0.1584	1
Correlation Coefficient Matrix of Output Indexes	ID.7	ID.8	ID.9	ID.10	ID.11	ID.12
ID.7	1	0.8349	0.6379	0.485	0.8349	0.7693
ID.8	0.8349	1	0.6198	0.3707	0.9408	0.9222
ID.9	0.6379	0.6198	1	0.9358	0.7155	0.522
ID.10	0.485	0.3707	0.9358	1	0.5078	0.274
ID.11	0.8349	0.9408	0.7155	0.5078	1	0.912
ID.12	0.7693	0.9222	0.522	0.274	0.912	1

3) Define a critical value M = 0.9.

If correlation coefficient of two indexes is greater than M, only one index is retained. Otherwise, both are retained. Table 3 below shows the result of screening index according to correlation matrix of Table 2 and critical value M = 0.9.

Table 3. Indexes of express enterprises after screening

Retained indexes	Deleted indexes	Correlation coefficient
ID.1	ID.2	0.9352
ID.8	ID.11	0.9408
	ID.12	0.9222
ID.10	ID.9	0.9358

5.1.3. *Screening index based on improved principal component analysis.* Indexes screened by correlation analysis include total staff number (ID.1), own vehicles (ID.3), enterprise management software (ID.4), city dot (ID.5), warehouse area (ID.6), defined as x_1, x_2, x_3, x_4, x_5 in order and output indexes are respectively total business revenue (ID.7), monthly processing order quantity (ID.8), delivery of cargoes (ID.10), defined as y_1, y_2, y_3 in order. Calculate express enterprises indexes by improved principal component analysis. Table 4 shows the corresponding relative contribution rates of input and output indexes.

Table 4. Relative contribution values of input and output indexes

Contribution values of input indexes		Contribution values of output indexes	
x_2	0.3091	y_1	0.4670
x_1	0.2897		
x_5	0.2871	y_2	0.4265
x_3	0.0698		
x_4	0.0443	y_3	0.1065

From the table above, cumulative contribution rate of variables x_2, x_1, x_5 achieves 88.59%, which means original five variables can be replaced with those three variables; cumulative contribution rate of variables y_1, y_2 achieves 89.35%, which means original three variables can be replaced with those two variables.

Therefore, after screening by improved principal component analysis, deleted indexes of express-delivery enterprise are enterprise management software (ID.4), city dot (ID.5), delivery of goods (ID.10). Remaining indexes are own carrier (ID.3), total staff (ID.1), warehouse area (ID.6), total business revenue (ID.7), monthly processing order quantity (ID.8).

5.2. **Index distribution of DEA evaluation model.** Through numerous indexes screened by correlation analysis and improved principal component analysis, final simplified indexes of each category of logistics enterprises are obtained. The relative quantity of indexes and DMUs of each category has met the requirement for "freedom degree" of DEA model. These indexes correspond with input, output and intermediate indexes of DEA model and the evaluation that applies the DEA methods is in the following.

Taking express industry as an example, we compare improved two-stage DEA model with traditional EDA model and traditional two-stage DEA model and demonstrate the effectiveness of the improved two-stage DEA model. Index distribution of three evaluation models is shown in Table 5.

Table 5. Index distribution of three evaluation models

<table>
<tr><td rowspan="5">Traditional DEA Evaluation Model</td><td rowspan="3">Input Indexes</td><td>ID.6</td><td rowspan="5">Traditional Two-stage DEA Evaluation Model</td><td rowspan="3">System Inputs</td><td>ID.6</td><td rowspan="5">Improved Two-stage DEA Evaluation Model</td><td rowspan="2">Original System Inputs</td><td>ID.6</td></tr>
<tr><td>ID.3</td><td>ID.3</td><td>ID.3</td></tr>
<tr><td>ID.1</td><td>ID.1</td><td>Shared Inputs of Two-stages</td><td>ID.1</td></tr>
<tr><td rowspan="2">Output Indexes</td><td>ID.8</td><td>Intermediate Products</td><td>ID.8</td><td>Intermediate Products</td><td>ID.8</td></tr>
<tr><td>ID.7</td><td>System Outputs</td><td>ID.1</td><td>System Outputs</td><td>ID.1</td></tr>
</table>

Method of index distribution of other industries is the same as that of express industry, which is not elaborated here.

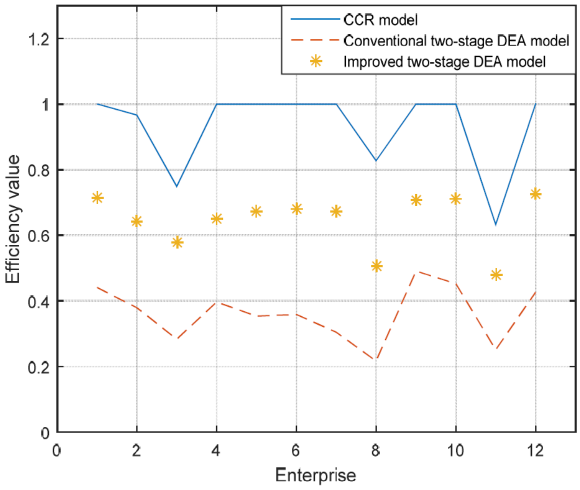
5.3. **Evaluation and analysis of operation efficiency of logistics enterprises in Xiamen.** According to the screening results of the above indicators, the indexes are allocated based on the traditional DEA model, the traditional two-stage DEA model, and the improved two-stage DEA model respectively. By using three models, calculate operation efficiency of different logistics enterprises in Xiamen. Evaluation results are shown in Figure 3.

From Figure 3, efficiency values of improved DEA model are mostly between those of traditional CCR model and those of traditional two-stage DEA model, explicating that improved DEA model that we used can effectively make up for the deficiency of higher results of CCR model and lower results of traditional two-stage DEA model.

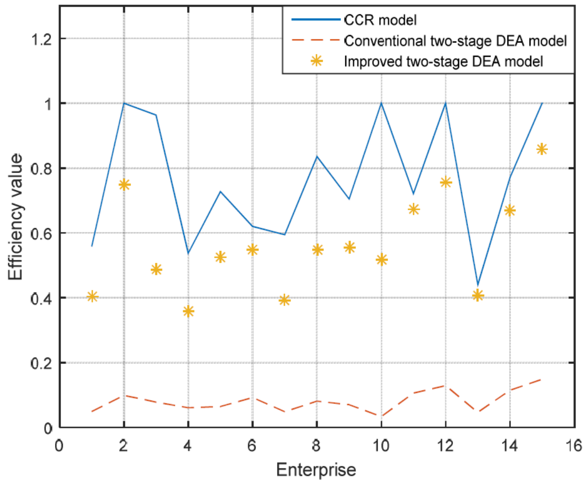
We regard average efficiency of each type of enterprises as overall efficiency of this type of enterprises, in order to reflect the overall development of urban logistics, overall efficiency of various types of logistics enterprises are needed to be compared and analyzed. The results are shown in Figure 3(f) and Table 6.

As can be seen from Figure 3(f) and Table 6, no matter which model is used in the calculation, evaluation result of express industry's operation is the best and that of warehousing industry's operation is comparatively the worst, while results of 3PL, freight forwarding, and transportation industry are between results of warehousing industry and express industry. It can be concluded that management level of warehousing industry is relatively inadequate, which means that it has a greater improvement in space. Under the benchmark effect of express industry, 3PL, freight forwarding and transportation industry should obtain advanced experience from operation mode of 3PL, and thus further improve operation standard.

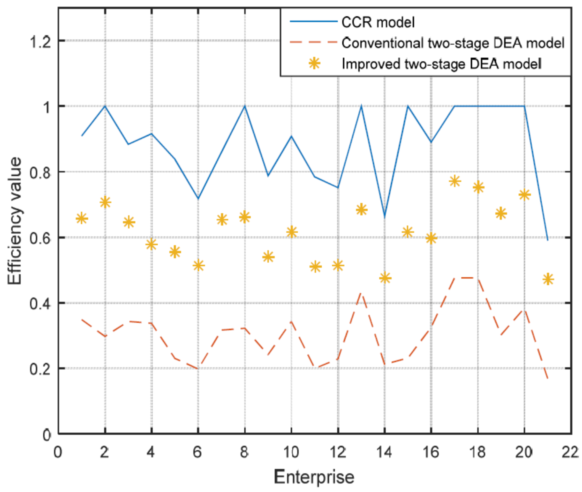
We compare overall efficiency ranking among 3PL, freight forwarding and transportation industry in different DEA models. Overall efficiency of 3PL in CCR model and improved two-stage DEA model is the best among the three. Only in the traditional two-stage DEA model, it ranks the second in the three, and under this model, overall efficiency value of 3PL is 0.3053, which is not far from 0.3284 of transportation industry that is ranked the second; therefore, although overall efficiency of 3PL is not always greater than the others in three models, it is optimal according to the ranking. Between freight forwarding and transportation industry, although overall efficiency of freight forwarding is placed the first in front of that of transportation industry in CCR model and improved two-stage model, overall efficiency of freight forwarding falls two ranks behind that of transportation industry and the difference is great; therefore, it is inappropriate to compare these two industries further.



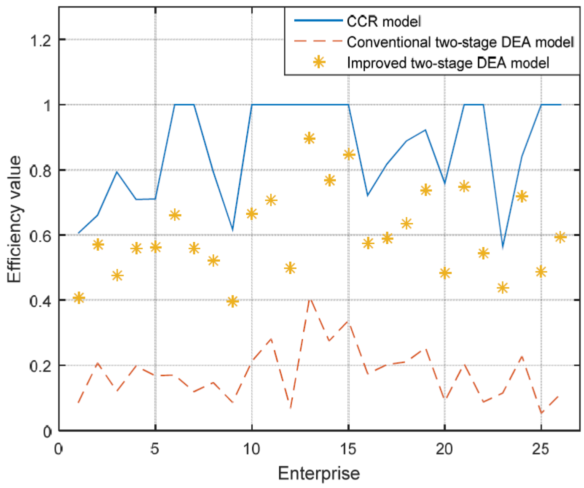
(a) The evaluation results of express industry



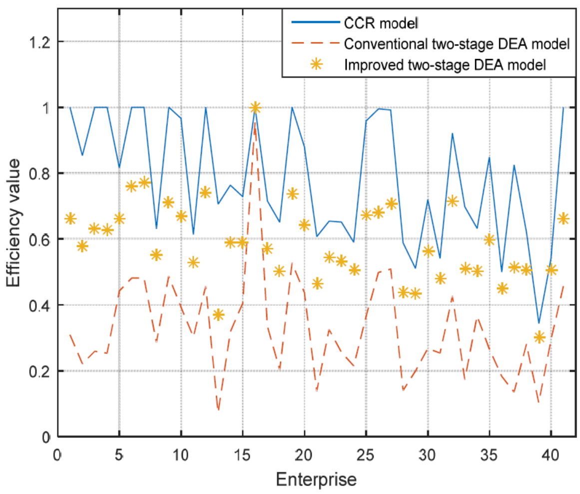
(b) The evaluation results of warehousing industry



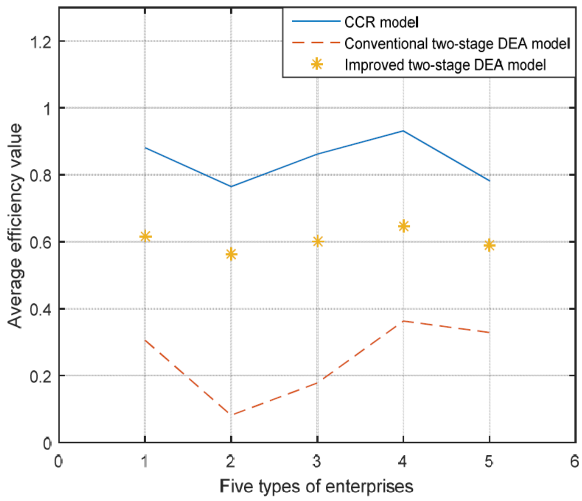
(c) The evaluation results of third party logistics industry



(d) The evaluation results of freight forwarding industry



(e) The evaluation results of transportation industry



(f) Efficiency mean of different types of enterprises

FIGURE 3. The evaluation results under three models

Table 6. Efficiency average of various categories of enterprises under three models

	CCR model		Traditional two-stage DEA model		Improved two-stage DEA model	
	Efficiency average	Rank	Efficiency average	Rank	Efficiency average	Rank
3PL	0.8809	2	0.3053	3	0.6151	2
Warehousing	0.7648	5	0.0811	5	0.5628	5
Freight forwarding	0.8616	3	0.1774	4	0.6013	3
Express	0.9312	1	0.3626	1	0.6447	1
Transportation	0.7819	4	0.3284	2	0.5892	4

In all, in the whole logistics industry, express industry and warehouse industry are respectively "improving" efficiency value and "pulling down" efficiency value, which means operation of express industry is the best and that of warehousing industry is the worst. 3PL, freight forwarding and transportation are between express and warehousing and 3PL is better than freight forwarding and transportation.

6. **Conclusion.** We evaluate operation efficiency of logistics enterprises by DEA method and two-stage DEA model, delete redundant indexes by correlation analysis and improved principal component analysis and allocate indexes appropriately. Following conclusions are summarized from the empirical analysis of five categories of logistics enterprises in Xiamen.

1) Index screening by adapting DEA method has great effect on evaluation results. The paper screens indexes with high correlation by correlation analysis, analyzes their impact degree on evaluation results by improved principal component analysis and deletes indexes with low impact degree. Therefore, information is retained to the greatest extent as the dimensions of indexes are declined.
2) Traditional DEA method fails to take internal structure of DMU into account, leading to higher efficiency value. Although traditional two-stage DEA model has opened the black box, structure problem of input and intermediate products has not been further discussed, leading to lower efficiency value. However, evaluation results of improved two-stage DEA model are between them, proving that improved two-stage DEA method is more effective.
3) Through efficiency evaluation of logistics enterprises classification, logistics efficiency difference among various types of enterprises or within the same type of enterprises is obtained, which benefits the scientific decision making of logistics network design and development plan of enterprises for themselves.

Two-stage DEA method we have adopted does not discuss frontier projection but reflects relative effectiveness of whole logistics enterprises by efficiency values. Final efficiency value is not applied to solving frontier projection. It is suggested that further studies should focus on how to solving front projection by improved two-stage DEA model we have used.

Data used in this paper have many indexes, but "freedom degree" is required in DEA method. How to minimize the number of indexes while ensuring integrity and objectivity of the information can be discussed in future studies.

Acknowledgment. The research is supported by Key Project of Science and Technology R&D Plan of China Railway Corporation (No. 2017X009-J) and the National Youth Foundation of China under Grants (No. 71603162).

REFERENCES

[1] A. Charnes, W. W. Cooper and E. Rhodes, Measuring the efficiency of decision-making units, *European Journal of Operational Research*, vol.2, no.6, pp.429-444, 1978.

[2] R. D. Banker, A. Charnes and W. W. Cooper, Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, vol.30, no.9, pp.1078-1092, 1984.

[3] P. Byrnes, R. Färe and S. Grosskopf, Measuring productive efficiency: An application to Illinois strip mines, *Management Science*, vol.30, no.6, pp.671-681, 1984.

[4] C. Kao, Some properties of Pareto efficiency under the framework of data envelopment analysis, *International Journal of Systems Science*, vol.26, no.9, pp.1549-1558, 1995.

[5] K. H. Choi and J. Y. Ahn, Measuring the efficiency of universities in Korea using dea and clustering approach, *ICIC Express Letters, Part B: Applications*, vol.5, no.6, pp.1673-1679, 2014.

[6] R. Färe and G. Whittaker, An intermediate input model of dairy production using complex survey data, *Journal of Agricultural Economics*, vol.46, no.2, pp.201-213, 2010.

[7] R. Färe and S. Grosskopf, Network DEA, *Socio-Economic Planning Sciences*, vol.34, no.1, pp.35-49, 2000.

[8] L. Castelli, R. Pesenti and W. Ukovich, A classification of DEA models when the internal structure of the decision making units is considered, *Annals of Operations Research*, vol.173, no.1, pp.207-235, 2010.

[9] W. D. Cook, L. Liang and J. Zhu, Measuring performance of two-stage network structures by DEA: A review and future perspective, *Omega*, vol.38, no.6, pp.423-430, 2010.

[10] C. Kao, Network data envelopment analysis: A review, *European Journal of Operational Research*, vol.239, no.1, pp.1-16, 2014.

[11] M. M. Yu and C. K. Fan, Measuring the cost effectiveness of multimode bus transit in the presence of accident risks, *Transportation Planning & Technology*, vol.29, no.5, pp.383-407, 2006.

[12] Y. Zha, Two-stage cooperation model with input freely distributed among the stages, *European Journal of Operational Research*, vol.205, no.2, pp.332-338, 2010.

[13] Y. Chen, J. Du, H. D. Sherman and J. Zhu, DEA model with shared resources and efficiency decomposition, *European Journal of Operational Research*, vol.207, no.1, pp.339-349, 2010.

[14] L. Liang, F. Yang, W. D. Cook and J. Zhu, DEA models for supply chain efficiency evaluation, *Annals of Operations Research*, vol.145, no.1, pp.35-49, 2006.

[15] Y. Chen, W. D. Cook, C. Kao and J. Zhu, Network DEA pitfalls: Divisional efficiency and frontier projection, *European Journal of Operational Research*, vol.226, no.3, pp.507-515, 2013.

[16] S. Lozano, E. Gutiérrez and P. Moreno, Network DEA approach to airports performance assessment considering undesirable outputs, *Applied Mathematical Modelling*, vol.37, no.4, pp.1665-1676, 2013.

[17] M. Maghbouli, A. Amirteimoori and S. Kordrostami, Two-stage network structures with undesirable outputs: A DEA based approach, *Measurement*, vol.48, no.1, pp.109-118, 2014.

[18] Y. Yu and Q. Shi, Two-stage DEA model with additional input in the second stage and part of intermediate products as final output, *Expert Systems with Applications*, vol.41, no.15, pp.6570-6574, 2014.

[19] J. Ma, A two-stage DEA model considering shared inputs and free intermediate measures, *Expert Systems with Applications*, vol.42, no.9, pp.4339-4347, 2015.

[20] W. D. Cook and M. Hababou, Sales performance measurement in bank branches, *Omega*, vol.29, pp.299-307, 2001.