Research on Effectiveness and Scientificalness of Logistics Based on PCA-DEA Approach: An Illustrative Example from Yangtze River Delta Port of China

Changbing Jiang, Chunhua Ju, Fei Ding

Abstract

Yangtze River Delta region in China is rich in port resources. "To prosper city according to ports" has become one of the strategies of promoting economy development. Therefore, the development of port logistics has a very important position in Yangtze River Delta region economy development. This paper sets up a new evaluation index system for port logistics efficiency and effectiveness evaluation for 16 municipal ports in Yangtze River Delta region, then an empirical study is carried out by using Principal Component Analysis (PCA) and Data Envelopment Analysis (DEA) approach. According to the double evaluation system of PCA and DEA, the port industrial clusters in Yangtze River Delta region is divided into four categories to measure the effectiveness and efficiency of port logistics infrastructure.

Keywords: Port Logistics, Principal Component Analysis, Data Envelopment Analysis

1. Introduction

Being one of the most developed areas in China, the Yangtze River Delta region possessed its GDP at 4.69 trillion in 2007, at a growth of 14.1%, accounting for 18.8% of the national economy. What’s more, its growth was much higher than the national average growth of 11.1%, accounting for one fifth of the total domestic logistics. Its unique geographical location and economic develop level has laid a good foundation for the local development of modern logistics industry [1]. The following years is a fast yet steady development period for national economy, also stands for a period that Yangtze River Delta will face in the new round of rapid development. Additionally, the logistics industry in the Yangtze River Delta will meet a lot of strategic opportunities of great-leap-forward development. The port in the Yangtze River Delta plays a key point in ameliorating local investment environment and promoting regional competitiveness as well as export-oriented economic development services. Port is vital in integrated logistics system, is the rally point where maximum of goods flows to in the transportation chain, assembling a large number of goods from inland and sea transport. In addition, port acts as an important service base and distribution center in international trade. Therefore, in the process of developing modern logistics and constructing logistical center, port plays a leading and key role in development of modern logistics, owning to its large scale of transportation capacity and the role in organizing logistics network. At the same time, developing port logistics is beneficial to expanding the port throughput by combined effect of cargos, which can add to the values of port enterprises, and comprehensively enhance the competitiveness of the port.

In recent years, with the rapid development of the port construction in Yangtze River Delta, each port lacked overall consideration under construction. They invested to expand their docks from their own economic interests, resulting in unclear division of labor, blind competition, overlapping investment and low-level redundant construction, which in turn led to the unreasonable structure of regional port layout. Meanwhile, the increasingly intense competition among neighboring ports prevents the carriers from choosing the correct port in accordance with
the reasonable cargo route, producing a large number of irrational transportation and high cost of port logistics. Therefore, port logistics system lacks overall coordination. The unreasonable layout of regional ports leads to lower international competitiveness of regional port cluster. Therefore, it is of great importance of the Yangtze River Delta port cluster to achieve an effective layout, rational allocation of resources and reasonable position of port functions. It helps to improve the international competitiveness of Yangtze River Delta port cluster [2].

This paper resorts to Principal Component Analysis (PCA) and Data Envelopment Analysis (DEA) approach to compare and analyze the logistics efficiency and effectiveness evaluation of 16 municipal ports in Yangtze River Delta region respectively, and construct a matrix using PCA scores as vertical axis and DEA scores as horizontal axis. Using this method, the Yangtze River Delta region is divided into four categories, and the developing directions of all categories are preliminary given.

2. Research methodology

2.1 Principal component analysis

Principal Component Analysis (PCA) was first proposed by Karl and Pearson in 1901, which at that time was applied to non-random variables. Later on it was promoted by Hotelling to the random vector in 1933. PCA is widely used to reduce number of variables under study to a few comprehensive and standardized variables, using dimension reduction techniques. Its principle is to reorganize the original, numerous and relevant variables into a new group of uncorrelated variables through linear combination, which is called principal components [3]. Generally, the first principal component displays the largest variation in the data. If the first new component fails to illustrate all original information, then take the second principal component into account in the premise that no repeated information is included. And in turn so on, until the selected principal components can include adequate information in evaluation.

The PCA method studies the inner structural relationship of index system. It translates several variables into a few independent variables, whose eigenvalues are larger than 1. The advantages lie in the fact that those principal components represent the inner structure of the indicators, and their weights are based on analysis of data, not subjective factors. So they possess a good objectivity. The consequent comprehensive variables reduce the cross-information, which is very beneficial for analysis and evaluation.

PCA is achieved through a set of well-defined steps as follows [4]:

(1) Normalize the indicator vectors. The indicators must be normalized and have same order to be used in PCA.
(2) Standardize the indicators. They are standardized through predefined mean and standard deviation for each indicator.
(3) Evaluate the correlation matrix. This matrix shows the values of linear correlation of the original indicators.
(4) Calculate eigenvalues, eigenvectors and proportion of the sample variance for all the principal components (new variables). Determine the number of principal components that account for the most percentage of the sample variance. The coefficients \( \lambda_i \) are retrieved from the eigenvectors for the respective principal components.
(5) Evaluate principal components and aggregated weights. Determine the values of principal components, and consequently PCA scores.

The formula for scoring the principal components:

\[
F_i = a_{1i}ZX_1 + a_{2i}ZX_2 + \cdots + a_{pi}ZX_p
\]  

In the formula, \( F_i \) is the score of each principal component, \( p \) is the number of original variables, \( ZX_p \) is normalized from original data \( X_p \), \( a_{1i}, a_{2i}, \ldots, a_{pi} \) (\( i=1, 2, \ldots, n \)) are the eigenvectors of the eigenvalues of the covariance matrix.

The formula for scoring and evaluating the comprehensive principal components:
Above all, $\lambda_i$ represents the eigenvalues corresponding to each principal component, while $F_n$ is identified as score of each principal component.

### 2.2 Data envelopment analysis

In recent decades, researches on measuring port efficiency and effectiveness have been on the stage, especially using a method called Data Envelopment Analysis. Oh and Park proposed using international competitiveness to measure productivity [5]. Roll and Hayuth employed DEA method to analyze the efficiency and productivity in international ports like Australia [6]. The most profound research recently includes the application of several Data Envelopment Analysis models and random boundary method in using small sample data to analyze container ports efficiency [7].

Data Envelopment Analysis (DEA) is a new field integrating operations research, management science and mathematical economics. DEA was developed by Charnes, Cooper, and Rhodes [8]. DEA measures technological efficiency by relating multiple inputs to multiple outputs of decision making unit (DMU). Based on the observed data, we can determine whether DMU is DEA effective, and compare decision making units. DEA generates a surface called the frontier that follows the peak performers and envelops the remainder. Thus the effectiveness of DMU can be drawn from the distance between DMU and the frontier. Also, we can conclude the reason of non-DEA efficient or weak DEA efficient, as well as its direction and extent to improve through projection. DEA can be used not only to judge whether DMU is located on the efficient production frontier, but also gain a lot of useful information. As a multi-index decision making method, DEA doesn’t need coefficient of each index, nor any subjective judgments to decide the weight beforehand. Therefore, DEA analysis reduces impact from subjective factors and the evaluation can better reflect the actual state of decision making units.

Suppose there are $n$ decision making units, each unit has $a$ inputs and $b$ outputs. We use $A_j$ and $B_j$ to identify the input vector and output vector of the $j$th unit. We mark $A_j=(a_{ij}, a_{2j}, \ldots, a_{mj})^T$, $B_j=(b_{1j}, b_{2j}, \ldots, b_{sj})^T$.

The DEA model formula is shown below:

\[
D : \quad \min \quad V_{D_\varepsilon} = y \\
\text{s.t.} \quad \sum_{j=1}^{a} a_{ij} x_j + s_i = a_{ij} y, i=1,2,\ldots,m \\
\sum_{j=1}^{b} b_{kj} x_j - t_k = b_{kj}, k=1,2,\ldots,s \\
\quad x_j \geq 0, j=1,2,\ldots,n; s_i \geq 0, i=1,2,\ldots,m; t_k \geq 0, k=1,2,\ldots,s
\]

(3)

Where

$s_i$: the slack variable of $i$th input indicators

$t_k$: the slack variable of $k$th output indicators

$y$: proportional variable

Introducing non-Archimedes infinitesimal quantity to solve the degenerate case in linear programming, we can get the new C$^2$R model:

\[
D_\varepsilon : \quad \min \quad V_{D_\varepsilon}(\varepsilon) = y - \varepsilon(\sum_{i=1}^{a} s_i + \sum_{k=1}^{b} t_k)
\]
Suppose the optimum solutions are $y^*$, $s_i^*$, $t_k^*$, some effective judgment rules will be concluded:

(1) $y^*=1$, $s_i^*=0$ ($i=1,2,\ldots,m$); $t_k^*=0$ ($k=1,2,\ldots,s$), then DMU $j_0$ is DEA effective, and its economic significance means the optimal innovative efficiency and constant return to scale simultaneously.

(2) $y^*=1$, $s_i^*$ and $t_k^*$ not equals zero, we can conclude that DMU $j_0$ is weak ineffective.

(3) $0<y^*<1$, then DMU is DEA inefficient. Its economic significance means it is not simultaneous to reach optimal innovative efficiency and constant return to scale.

3. Evaluation system of port logistics

3.1 Set up indicators

According to the principle and feature of DEA, we should try to use fewer inputs in return to more outputs in the process of evaluation. It’s important to guarantee the evaluation indicators to be precisely used. Also, the direct relationship of inputs and outputs should be taken into consideration. When selecting specific indicators, we can cut off some correlate indicators through PCA or fuzzy clustering analysis.

Port logistics system is an important part in logistics service chain. It aims at satisfying consumers’ requirements. It uses related information and logistics activities to achieve effective planning, implementation and control. In this way, it realizes high efficiency of goods, high benefit of loading and unloading, high responding speed and value-added services. The effectiveness of port logistics refers to the effectiveness of port operation and regional economic output in the condition that efforts have done to port logistics infrastructure. The effectiveness of port logistics involves 4 aspects. (1) The priority is to have a large throughput capacity. It means not only the annual throughput capacity of cargos but also throughput capacity of container goods should reach a certain degree. (2) Infrastructure improvement. Water depth in approach channel and at berth can meet the expectations, and port facilities meet the demand of developing modern logistics. (3) A superior developing environment of ports. An advanced port should be equipped with excellent natural conditions, location advantages, economic prosperity of hinterland and developed adjacent industry. (4) Potential of sustainable development, including steady increasing in throughput and growing regional economic indicators [9][10].

The output indicators used: cargo throughput ($O_1$), total imports and exports ($O_2$), growth rate of gross regional product(GRP)($O_3$), annual growth rate of cargo throughput ($O_4$). Following are inputs: gross regional product(GRP) ($I_1$), fixed investment in logistics ($I_2$), number of berths ($I_3$), number of motor boats ($I_4$), number of barges ($I_5$). The indicator system is provided in Table 1 [11].
Table 1. The input and output indicators system of port logistics

<table>
<thead>
<tr>
<th>First-grade factor</th>
<th>Second-grade factor</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs (Oi)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Port operating condition</td>
<td>$O_1$: Cargo throughput</td>
<td>million ton</td>
</tr>
<tr>
<td></td>
<td>$O_2$: Total imports and exports</td>
<td>billion USD</td>
</tr>
<tr>
<td>Potential for development of port logistics</td>
<td>$O_3$: Growth rate of gross regional product (GRP)</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>$O_4$: Annual growth rate of cargo throughput</td>
<td>%</td>
</tr>
<tr>
<td>Outputs (Ii)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developing environment port logistics</td>
<td>$I_1$: Gross regional product (GRP)</td>
<td>billion yuan</td>
</tr>
<tr>
<td></td>
<td>$I_2$: Fixed investment in logistics</td>
<td>million yuan</td>
</tr>
<tr>
<td>Port infrastructure</td>
<td>$I_3$: Number of berths</td>
<td>unit</td>
</tr>
<tr>
<td></td>
<td>$I_4$: Number of motor boats</td>
<td>unit</td>
</tr>
<tr>
<td></td>
<td>$I_5$: Number of barges</td>
<td>unit</td>
</tr>
</tbody>
</table>

3.2 Data source

The paper is based on the Yangtze River Delta region and 16 municipal cities, we focused on their effectiveness of inputs and outputs. All the statistics come from Statistical Yearbook of Yangtze River and Pearl River Delta and Hong Kong, Macao, Taiwan (2010) [12]. See Table 2.

4. The empirical analysis

4.1 Efficiency of port logistics

In order to analyze the operating efficiency of port logistics, we use the statistical package for social sciences (SPSS) software, where we use the Principal Component Analysis into application. We focus on 16 cities located in the Yangtze River Delta region, and select 4 outputs as indicators. First of all, the original data are standardized to get the eigenvalues of correlation matrix.

Through the principal component analysis of major economic indicators in each port city, we can get Table 3, which is informative. The first two principal components both have the eigenvalue beyond 1, and the cumulative variance of the two reaches 84.007%, which demonstrates the inclusion of most information. We diminish the 4 indicators into 2. According to the initial factor loading matrix analysis of those 2 principal components, we can rename the 2 principal components. The first principal component stems mostly from throughput of cargos as well as total imports and exports, with the initial factor loading of 0.885 and 0.915 respectively. Thus we can define the first principal component as the operation condition of the port. Indicators that contribute mostly to the second principal component is growth rate of gross regional product and annual growth rate of cargo throughput, whose initial factor loading are 0.762 and 0.811. The second principal component can be entitled potential of port logistics.

We not only solve the problem of massive indicators, but also get 2 principal components that can almost reflect the developing level of port logistics. They are of practical significance. The eigenvalues of the 2 principal components are $\lambda_1=1.905$, $\lambda_2=1.455$. Since the variables are standardized, we can eliminate the influence of different dimensions of variables and different order of magnitude. The expressions of those 2 principal components are:

$F_1=0.641*O_1+0.663*O_2-0.306*O_3-0.235*O_4$

$F_2=0.303*O_1+0.238*O_2+0.632*O_3+0.672*O_4$

Then, using the scores and weight coefficients of each principal component, we can calculate the eventually score, $F$, which can measure the competitiveness of port logistics. We derive it:

$$F = \frac{\lambda_1}{\lambda_1+\lambda_2} F_1 + \frac{\lambda_2}{\lambda_1+\lambda_2} F_2$$
Above all, \( \frac{\lambda_1}{\lambda_1 + \lambda_2} \) refers to weight coefficient, \( F_i \) stands for score of the \( i \)th principal component, and thus we can quantify the score \( F \):

\[
F = 0.495*O_1 + 0.479*O_2 + 0.100*O_3 + 0.158*O_4
\]

The comprehensive scores can be seen in Table 4 named “PCA score”. From the column “PCA score”, we can infer that only 5 cities score more than 0, and others falling behind the average level. The leading 3 cities are Shanghai, Suzhou and Ningbo. Port logistics efficiency of cities are between big gaps.

**Table 2.** Input and output logistics indicator data of Yangtze River Delta region

<table>
<thead>
<tr>
<th>DMU</th>
<th>Cities</th>
<th>Outputs</th>
<th>Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Shanghai</td>
<td>561.4/5</td>
<td>2829.73</td>
</tr>
<tr>
<td>2</td>
<td>Nanjing</td>
<td>116.1/2</td>
<td>362</td>
</tr>
<tr>
<td>3</td>
<td>Suzhou</td>
<td>232.1/5</td>
<td>2117.96</td>
</tr>
<tr>
<td>4</td>
<td>Wuxi</td>
<td>146.4/9</td>
<td>511.46</td>
</tr>
<tr>
<td>5</td>
<td>Changzhou</td>
<td>65.61</td>
<td>132.26</td>
</tr>
<tr>
<td>6</td>
<td>Zhenjiang</td>
<td>88.49</td>
<td>63.05</td>
</tr>
<tr>
<td>7</td>
<td>Nantong</td>
<td>161.3/3</td>
<td>127.76</td>
</tr>
<tr>
<td>8</td>
<td>Yangzhou</td>
<td>55.49</td>
<td>44.59</td>
</tr>
<tr>
<td>9</td>
<td>Taizhou</td>
<td>61.46</td>
<td>39.56</td>
</tr>
<tr>
<td>10</td>
<td>Hangzhou</td>
<td>79.08</td>
<td>434.3</td>
</tr>
<tr>
<td>11</td>
<td>Ningbo</td>
<td>345.1/9</td>
<td>564.99</td>
</tr>
<tr>
<td>12</td>
<td>Jiaxing</td>
<td>24.18</td>
<td>160.62</td>
</tr>
<tr>
<td>13</td>
<td>Huzhou</td>
<td>42.04</td>
<td>42.65</td>
</tr>
<tr>
<td>14</td>
<td>Shaoxing</td>
<td>13.42</td>
<td>192.95</td>
</tr>
<tr>
<td>15</td>
<td>Zhoushan</td>
<td>128.1/8</td>
<td>40.76</td>
</tr>
<tr>
<td>16</td>
<td>Taozhu</td>
<td>35.07</td>
<td>110.93</td>
</tr>
</tbody>
</table>

**Table 3.** The eigenvalues and contribution rate of PCA

<table>
<thead>
<tr>
<th>Component</th>
<th>Initial Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>1</td>
<td>1.905</td>
</tr>
<tr>
<td>2</td>
<td>1.455</td>
</tr>
<tr>
<td>3</td>
<td>0.481</td>
</tr>
<tr>
<td>4</td>
<td>0.158</td>
</tr>
</tbody>
</table>
4.2 Effectiveness of port logistics

From the data gathered in Table 2, together with the use of standard DEA models, we can come up with the overall effectiveness of the 16 cities. See DEA scores in Table 4. In the perspective of overall effectiveness, among the 16 cities in Yangtze River Delta region, only 9 cities are overall effective. They are Suzhou, Ningbo, Nantong, Zhoushan, Zhenjiang, Yangzhou, Taozhu, Huzhou and Shaoxing. The rest 7 cities including Shanghai, Nanjing, Hangzhou are overall ineffective. In contrast to its large numbers of inputs in operation elements, these cities are relatively ineffective in operating port logistics.

![Table 4. The comprehensive score of each city](image)

4.3 PCA-DEA analysis

We can get the PCA-DEA evaluation model of 16 cities in the Yangtze River Delta region through PAC and DEA analysis. In order to realize it, we regard PCA and DEA scores as the measurement of port logistics’ efficiency and effectiveness. Respectively, we use PCA scores and DEA scores as horizontal and vertical axis to establish a two-dimensional coordinate. The range of horizontal axis is \([-1.5, 3.0]\), 0 is selected as the threshold value. As for vertical axis, the range and boundary fall in \([0.3, 1]\) and 0.7. From the evaluation model we can see the distribution of cities in Figure 1.

Observe the city distribution in Figure 1, we can divide the 16 cities into 4 categories.

1. Efficient port cluster
   In this cluster, both PCA score and DEA score are high in the region (PCA score >= 0 and DEA score >= 0.7), including Shanghai, Suzhou, Ningbo and Nantong. These cities not only possess effective port logistics, but have high efficiency, which result in their high potential in developing port logistics.

2. Potential port cluster
   In this cluster, PCA score is low but DEA score is high in the region (PCA score < 0 and DEA score >= 0.7), including Shaoxing, Huzhou, Taozhu, Yangzhou, Taizhou, Zhenjiang, Zhoushan, Jiaxing and Changzhou. The port logistics of these cities are relatively effective, but inefficient, their development for port logistics has great potential.

3. General port cluster
   In this cluster, PCA score is high but DEA score is low in the region (PCA score >= 0 and DEA score < 0.7), including Wuxi. Port logistics in this cluster is relatively ineffective, but highly efficient. The development of port logistics needs to optimize industrial factors.

4. Inefficient port cluster
   In this cluster, both PCA score and DEA score are low in the region (PCA score < 0 and DEA score < 0.7), including Nanjing and Hangzhou. This indicates that the regional port logistics is neither effective nor efficient. To develop better logistics, they should not only optimize industrial factors, but also increase the infrastructure investment.
The paper integrates principal component analysis (PCA) and data envelopment analysis (DEA), establishes the combined evaluation framework. Through 4 outputs and 5 inputs, we can measure the effectiveness and efficiency of port logistics infrastructure. We come up to the following results:

1. When evaluating operation condition of port logistics, we should not only consider the efficiency of port logistics, but also consider the effect of logistics infrastructure.

2. This paper presents a dual evaluation system on efficiency and effectiveness of regional port. Port of the Yangtze River Delta region is divided into four clusters, including efficient port cluster, potential port cluster, general port cluster and inefficient port cluster.

3. Through the empirical research of the 16 city ports’ logistics infrastructure in Yangtze River Delta economic region, we draw the conclusion that the combined evaluation method can get reasonable, objective and scientific results.

6. Acknowledgment

This research is supported by Science Foundation of Ministry of Education of China (11YJC630081); Science and Technology Project of Zhejiang Province (2012C23016, 2012C25100); Zhejiang Natural Science Foundation (Y6090015) and the Contemporary Business and Trade Research Center of Zhejiang Gongshang University which is the Key Research Institute of Social Sciences and Humanities Ministry of Education (12JDSM11YB).
7. References


