Comprehensive Evaluation Model on the Value of Investment in Colleges and Universities

YUTIAN LIU\textsuperscript{1}, a*, SIYI WU\textsuperscript{2,b} and CHAO HAN\textsuperscript{3,c}

\textsuperscript{1}School of Economics, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, China, 200433
\textsuperscript{2} School of Statistics and Management, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, China, 200433
\textsuperscript{3} School of Information Management and Engineering, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, China, 200433

\textsuperscript{a}2013111831@live.sufe.edu.cn, \textsuperscript{b} wusiyi1996@126.com, \textsuperscript{c} 1647819676@qq.com

Keywords: comprehensive evaluation; fuzzy mathematics; analytic hierarchy process

Abstract. The paper conducted a comprehensive scoring and ranking on the public welfare investment value of all the universities and colleges in America through establishing the comprehensive evaluation model. Firstly, those universities and colleges with insufficient information disclosure were removed from the data and the multiple linear regression method was used to supplement some remaining missing data. Secondly, 15 relevant variables were selected and then combined with multi-level fuzzy comprehensive evaluation (MFCE) model and analytic hierarchy process (AHP) model to obtain the composite index of all the American universities and colleges. Subsequently, American universities and colleges were ranked in accordance with the index and those with top rankings deserved more investment. Finally, the sensitivity analysis was conducted on the scoring and ranking of American universities and colleges.

1. Introduction

It has been a kind of common problem in our social life that rather limited resources can be used in conducting the comprehensive analysis on the relevant aspects of target objects and obtaining the ranking of the target groups to further guide our choices. The act of ranking universities and colleges is familiar to us. Students make use of the ranking to choose which school to attend [1], while scientific researchers use it to select the suitable research environment [2]. The public welfare investment institutions may need a kind of evaluation method for investigating the value of investment in universities and colleges to help them find out which universities and colleges are better worth their public welfare investment in promoting education [3,4,5].

With the focus on the public welfare investment and a model established to comprehensively evaluate the value of investment in universities and colleges, this paper aimed to help those investment groups or individuals working on public benefits make the sensible investment allocations in American universities and colleges so that their investment not only went to those students who needed it most but also got the optimal investment return.

2. Notation Description

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCM2 (E_1)</td>
<td>Whether the school are on Heightened Cash Monitoring 2 (1 or 0)</td>
</tr>
<tr>
<td>PREDDEG (E_2)</td>
<td>Predominant degree awarded (0,1,2,3)</td>
</tr>
<tr>
<td>DEG_DIST (E_3)</td>
<td>Distribution index of the degrees awarded</td>
</tr>
<tr>
<td>DISTANCEONLY (E_4)</td>
<td>Flag for distance-education-only education (1 or 0)</td>
</tr>
</tbody>
</table>

\textsuperscript{*} Work supported by Shanghai University of Finance and Economics

© 2016. The authors - Published by Atlantis Press
UGDS (E₄)  Enrollment of undergraduate degree-seeking students
PPTUG_EF (E₆)  Share of undergraduate students who are part-time
NPT4 (E₇)  The average annual total cost of attendance
PCTPELL (E₈)  Percentage of undergraduates who receive a Pell Grant
PCTFLOAN (E₉)  Percentage of undergraduate students receiving a federal student loan
GRAD_DEBT (E₁₀)  Median debt of completers expressed in 10-year monthly payments
POOLED_SUPP (E₁₁)  Wide sense of completion rate.
LOCAL (E₁₂)  Locale of institution (1.5,1,0.5,0,-0.5,-1,-1.5,-2,-2.5)
MD_EARN (E₁₃)  Median earnings of students working and not enrolled 10 years after entry
GT_25K (E₁₄)  Share of students earning over $25,000/year 6 years after entry
RPY_3YR (E₁₅)  Three-year repayment rate
UG25abv  Percentage of undergraduates aged 25 and above
M₁  Financial demand of schools
M₂  Financial demand of students
T₁  Potential for effective use
T₂  Return on investment
S  Composite index

3. Data Analysis

3.1 Data Classification

The purpose of public welfare investment in universities and colleges is mainly to improve the quality of students’ learning and life. Universities and colleges as well as students that have higher demands for capital and could make better use of the investment deserve more investment. Therefore, this paper investigates the value of investment in universities and colleges from two aspects: potential for effective use and return of investment. Moreover, the potential for effective use is evaluated by two aspects: financial demand of schools and financial demand of students.

This paper selected the information data of 2,936 universities and colleges in America from the U.S. National Center on Education Statistics[6] and the College Scorecard data set[7] and classified 15 variables (E₁—E₁₅) relevant to the public welfare investment as shown in Fig.1.

E₅ represents the degree of distribution equilibrium about the academic degrees granted by universities and colleges in three different categories of subjects including Natural Science and Engineering, Social Science and Art, and Others. The larger index means that the subject distribution of the university is more imbalanced.

E₁₂ represents the type of the schools’ locations and its value is shown in Table 2.
Table 2: Value of E_{12}

<table>
<thead>
<tr>
<th>Location</th>
<th>Value</th>
<th>Location</th>
<th>Value</th>
<th>Location</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large city</td>
<td>1.5</td>
<td>Midsize city</td>
<td>1</td>
<td>Small city</td>
<td>0.5</td>
</tr>
<tr>
<td>Large suburb</td>
<td>1</td>
<td>Midsize suburb</td>
<td>0.5</td>
<td>Small suburb</td>
<td>0</td>
</tr>
<tr>
<td>Fringe town</td>
<td>0</td>
<td>Distant town</td>
<td>-0.5</td>
<td>Remote town</td>
<td>-1</td>
</tr>
<tr>
<td>Fringe rural</td>
<td>-1.5</td>
<td>Distant rural</td>
<td>-2</td>
<td>Remote rural</td>
<td>-2.5</td>
</tr>
</tbody>
</table>

3.2 Data Processing

After determining the above 15 variables, we find that some data values are missing or censored in some schools. Schools with insufficient information disclosure tend to have higher investment risks. To ensure the accuracy of the final analysis, the schools the missing value of which exceeds 3 among 15 variables are removed and 2,761 universities and colleges are left as candidate schools finally. Of these candidate schools, 344 schools have the problem of missing data: most schools miss one data. For schools missing more than one data, the situation is divided into the following five types:

a) Both E_{10} and E_{13} are missing or censored.

b) Both E_{10} and E_{14} are missing or censored.

c) Both E_{10} and E_{15} are missing or censored.

d) Both E_{13} and E_{14} are missing or censored.

e) E_{10}, E_{13} and E_{14} are all missing or censored.

Considering the missing data, this paper adopts the STATA software to conduct the multiple linear regression analysis and makes use of the complete data information of schools to obtain the multiple linear regression equations for these missing variables. Subsequently, the estimated value obtained from these equations is used to replace the missing value and is treated as the data of the school. The actual regression result is shown in Table 3.

It can be seen that all the regression equations are significant and that most variables are of statistical significance at 1%. This means that the regression is effective. Meanwhile, it is also discovered: the value of the goodness of fit R-square is not up to 0.9, but the value is generally larger. In the model which integrates so many variables for the evaluation of universities and colleges, such approximate substitution method is acceptable.

Moreover, the original data of E_{5} and E_{7} vary greatly between universities and colleges, so the difference between the maximum and the minimum can reach a thousand times. However, while considering the investment in universities and colleges, the final ranking of a school should not be determined by the number of students or high tuition. Therefore, the logarithm is taken based on the raw value of these two variables and then substituted into the comprehensive evaluation model.

4. Developing the Comprehensive Evaluation Model

4.1 Multi-level Fuzzy Comprehensive Evaluation (MFCE) Model

MFCE is the fuzzy mathematical method which comprehensively evaluates different aspects of a thing in the multi-factor participation decision-making process \[8\]. This paper integrates different indexes of 15 variables in different schools into a final composite index. Therefore, MFCE Model is selected and used to comprehensively evaluate the value of the investment in universities and colleges.

4.1.1 Factors Grading.

As shown in Fig. 1, factors are graded. The third-grade factor is classified into M_1 = (E_1, E_2, E_3, E_4, E_5), M_2 = (E_6, E_7, E_8, E_9, E_{10}), and T_2 = (E_{11}, E_{12}, E_{13}, E_{14}, E_{15}); and the second-grade factor is classified into T_1 = (M_1, M_2), and the first-grade factor is S = (T_1, T_2).
### Table 3: OLS Regression

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E_{10}</strong></td>
<td>4.61697</td>
<td>0.0000051</td>
<td>0.000214</td>
<td>-0.00020</td>
<td>21</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td><strong>E_{13}</strong></td>
<td><strong>(0.001)</strong>*</td>
<td><strong>(0.017)</strong></td>
<td><strong>(0.000)</strong>**</td>
<td><strong>(0.000)</strong>**</td>
<td>8.29e-06</td>
<td>(0.000)**</td>
<td>-1.72e-06</td>
<td>(0.000)**</td>
<td>1.637537</td>
</tr>
<tr>
<td><strong>E_{14}</strong></td>
<td>43.6770</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E_{15}</strong></td>
<td>-55.3312</td>
<td>0.00063</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E_{5}</strong></td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E_{11}</strong></td>
<td>2.5705</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E_{6}</strong></td>
<td>0.00221</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E_{7}</strong></td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E_{8}</strong></td>
<td>-94.828</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E_{9}</strong></td>
<td>252.547</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UG25abv</strong></td>
<td>-26.5579</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>2415</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>2415</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R-Square</strong></td>
<td>0.7013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F Statistic</strong></td>
<td>564.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prob. &gt; F</strong></td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

+ significant at 10%; * significant at 5%; ** significant at 1%
4.1.2 Developing the Single Factor Evaluation Matrix

Firstly, the membership functions are established for E1-E15. E3 represents the subject distribution equilibrium degree of schools. It is generally thought that: the more balanced the subject distribution is, the larger academic capital investment it will require. The value of E3 is inversely proportional to the distribution equilibrium degree \(0 \leq E_3 \leq 1\). Therefore, the membership function of factor E3 is as follows:

\[
C(E_3) = \begin{cases} 
1, & E_3 = 0 \\
1 - E_3, & 0 < E_3 < 1 \\
0, & E_3 = 1
\end{cases}
\]

(1)

The larger the values of the remaining factors are, the more it deserves the investment. Therefore, 14 other membership functions are

\[
C(E_i) = \begin{cases} 
0, & E_i = n \\
\frac{E_i - n}{m - n}, & n < E_i < m, \quad i \neq 3 \\
1, & E_i = m
\end{cases}
\]

(2)

Where, m is the maximum value of all the schools factor Ei, and n is the minimum value.

The value of 15 factors in 2,761 schools is substituted into membership functions and MATLAB is used for calculation. Subsequently, the corresponding single factor evaluation matrix of M1, M2 and T2 is as follows:

\[
\begin{bmatrix}
0 & \cdots & 0 & 1 & 0.8246 & 1 & \cdots & 0.5 & 0.4286 & 0 & 0.55112 & \cdots & 0.246162 \\
0.8246 & \cdots & 0.4286 & 0.626032 & \cdots & 0.678822 & \cdots & 0.437163 & \cdots & 0.7248 & \cdots & 0.875 \\
0.55112 & \cdots & 0.246162 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}_{5 \times 2761}
\]

\[
\begin{bmatrix}
0.0622 & \cdots & 0.7062 & 0.260898 & \cdots & 0.3622 & \cdots & 0.3622 & \cdots & 0.3622 & \cdots & 0.3622 \\
0.7115 & \cdots & 0.4297 & 0.19 & \cdots & 0.483333 & \cdots & 0.483333 & \cdots & 0.483333 & \cdots & 0.483333 \\
0.8204 & \cdots & 0.7248 & 0.437163 & \cdots & 0.813113 & \cdots & 0.813113 & \cdots & 0.813113 & \cdots & 0.813113
\end{bmatrix}_{5 \times 2761}
\]

\[
\begin{bmatrix}
0.5 \cdots 0.246162 \\
0.5 \cdots 0.7248 \\
0.5 \cdots 0.87756 \\
0.5 \cdots 0.87756 \\
0.5 \cdots 0.87756
\end{bmatrix}_{5 \times 2761}
\]

4.2 Analytic Hierarchy Process (AHP) Model

AHP is the important operational research approach which transforms the qualitative analysis into the quantitative analysis [\textsuperscript{9}]. This paper adopts AHP Model to determine the weight vector of comprehensive evaluation.

4.2.1 Developing the Pairwise Comparison Matrix

Pairwise comparison matrix \(D=\left(d_{ij}\right)_{n \times n}\), where \(d_{ij}\) represents the relative importance of factor \(F_i\) over factor \(F_j\) and satisfied the following conditions:

\[
d_{ij} > 0, \quad d_{ij} = 1/d_{ji}, \quad d_{ii} = 1, \quad i = 1, 2, \ldots
\]

(4)

The corresponding pairwise comparison matrix of M1, M2, T2, T1, and S is obtained:

\[
D_1 = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 \\
2 & 7 & 4 & 2 & 9 \\
5 & 1 & 2 & 7 & 2 \\
7 & 2 & 1 & 2 & 1 \\
4 & 1 & 2 & 1 & 3
\end{pmatrix}, \quad D_2 = \begin{pmatrix}
1 & 1 & 1 & 1 \\
1 & 2 & 2 & 2 \\
2 & 1 & 1 & 1 \\
2 & 1 & 1 & 1 \\
3 & 1 & 1 & 1
\end{pmatrix}, \quad D_3 = \begin{pmatrix}
1 & 5 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
5 & 1 & 9 & 5 & 7 \\
2 & 9 & 1 & 2 & 1 \\
1 & 5 & 1 & 1 & 1
\end{pmatrix}, \quad D_4 = D_5 = \begin{pmatrix}
1 & 1 \\
1 & 1
\end{pmatrix}
\]
4.2.2 Consistency Test and the Weight vector

Calculate the largest Eigen value and its eigenvector $A_i$ of each matrix $D_i$, and then we get the consistency ratio $CR_i = \frac{\lambda_i - n}{RI(n-1)}$, which is defined by AHP theory’s founder Satty. If $CR < 0.1$, it can be considered that the pairwise comparison matrix has satisfactory consistency and its Eigen vector $A_i$ is the weight vector. Otherwise, it is necessary to modify the pairwise comparison matrix. In the study, $n=5$ and the value of RI should be given 1.12 according to Satty’s result. $D_4$ and $D_5$ are obviously consistent matrix. It is obtained through calculation that $CR_1 = \frac{95}{1591}$, $CR_2 = \frac{75}{1693}$, $CR_3 = \frac{79}{2591}$, $CR_i < 0.1$ (i=1,2,3). Therefore, the weight vector is obtained as follows: $A_1 = \begin{pmatrix} 485 \\ 799 \\ 300 \\ 251 \\ 605 \end{pmatrix}$, $A_2 = \begin{pmatrix} 248 \\ 1050 \\ 1050 \\ 1050 \\ 1147 \end{pmatrix}$, $A_3 = \begin{pmatrix} 842 \\ 569 \\ 1647 \\ 5052 \\ 1647 \end{pmatrix}$, $A_4 = A_5 = (1 \ 1)$

4.3 Combine MFCE with AHP and Obtain Composite Index

MATLAB is used for calculation based on Eq. 3 and the following is obtained:

$R_4 = \begin{pmatrix} 0.022137 & \cdots & 0.008784 \\ 0.026935 & \cdots & 0.025715 \end{pmatrix}$, $R_5 = \begin{pmatrix} 0.025036 & \cdots & 0.017601 \\ 0.008973 & \cdots & 0.224435 \end{pmatrix}$

$S = A_5 \circ R_5 = (0.034008676 \ 0.043234387 \ 0.034039664 \ 0.040044009)$

To make the final composite index more visual, each component of $S$ is multiplied by 1697.792869 to make the maximum value reach 100 and obtain the final composite index distribution of all the schools as shown in Fig. 2 in this paper. The vertical coordinate of each data point is represented by the number of schools in the score interval of each score.

5. Sensitivity Analysis

5.1 Sensitivity Analysis on Pairwise Comparison Matrix

$D_5$ is changed into $D_5 = \begin{pmatrix} 1 & \frac{1}{3} \\ \frac{1}{3} & 1 \end{pmatrix}$ for analyzing the sensitivity of school scoring and ranking in the pairwise comparison matrix. The result is shown in Fig. 3 and Table 4.

Maximum variation rate of grade 50.74%
Percentage under the variation rate 20% 72.86%
Maximum variation rate of rank 24.19%
Percentage under the variation rate 20% 99.46%
It is obtained through analysis that the scoring and ranking for most of schools is controlled within 20% after changing the final pairwise comparison matrix. In other words, this comprehensive evaluation model is not highly sensitive to pairwise comparison matrix.

5.2 Sensitivity Analysis on Membership Function

The membership functions of E_7, E_9 and E_{10} are successively and constantly changed into

\[
C(E_i) = \begin{cases} 
1, & E_i = n \\
\frac{E_i - n}{m - n}, & n < E_i < m, \quad i = 7, 9, 10 \\
0, & E_i = m 
\end{cases}
\]  

This is to study the sensitivity of school scoring and ranking in the membership function. The result is shown in Fig. 4 and Table 5.

Figure 4: Sensitivity Analysis on Membership Function

<table>
<thead>
<tr>
<th>Changing one Membership Function</th>
<th>Maximum variation rate of grade</th>
<th>43.30%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage under the variation rate 20%</td>
<td>98.70%</td>
</tr>
<tr>
<td></td>
<td>Maximum variation rate of rank</td>
<td>23.32%</td>
</tr>
<tr>
<td></td>
<td>Percentage under the variation rate 20%</td>
<td>99.60%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changing two Membership Function</th>
<th>Maximum variation rate of grade</th>
<th>72.09%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage under the variation rate 20%</td>
<td>90.01%</td>
</tr>
<tr>
<td></td>
<td>Maximum variation rate of rank</td>
<td>32.96%</td>
</tr>
<tr>
<td></td>
<td>Percentage under the variation rate 20%</td>
<td>95.95%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changing three Membership Function</th>
<th>Maximum variation rate of grade</th>
<th>106.94%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage under the variation rate 20%</td>
<td>63.66%</td>
</tr>
<tr>
<td></td>
<td>Maximum variation rate of rank</td>
<td>55.16%</td>
</tr>
<tr>
<td></td>
<td>Percentage under the variation rate 20%</td>
<td>62.58%</td>
</tr>
</tbody>
</table>

In conclusion, it is found in the above sensitivity analysis that: the variation range of school scoring and ranking is below 20% when most of schools change the pairwise comparison matrix or the number of membership functions is less than 3. It can be seen in the graphic presentation that those schools with top ranking are less sensitive to pairwise comparison matrix and membership functions.

6. Conclusion

Through establishing the comprehensive evaluation model, this paper conducted the comprehensive scoring and ranking on the value of public welfare investment in American universities and colleges. Moreover, the information data of 2,936 schools collected from the U.S. National Center on Education Statistics and the College Scorecard data set were sorted out for analysis. 15 variables ranging from E_1 to E_{15} were selected as investigation factors to comprehensively evaluate the value of public welfare investment in American universities and colleges from two aspects: potential for effective use and return of investment.

Firstly, given that schools with insufficient information disclosure had relatively higher investment risks, some schools were removed from the data because of relatively more missing data, and 2,761 universities and colleges were selected as candidate schools for model evaluation.
Meanwhile, the multiple linear regression method was used to replace a minority of missing data with the regression estimation value.

Secondly, single factor evaluation matrix was obtained through multi-level fuzzy comprehensive evaluation (MFCE) model, and weight vector was obtained through analytic hierarchy process (AHP) model. Subsequently, MFCE and AHP were combined to obtain the composite index of all the 2,761 universities and colleges in America. It was found through analysis that the index basically complied with the normal distribution.

Finally, the sensitivity analysis was conducted on the scoring and ranking of universities and colleges and the analysis revealed that most of schools were insensitive to pairwise comparison matrix and membership functions. However, when the number of the altered membership functions was relatively large, great changes were found in the school scoring and ranking. In addition, schools with top ranking were generally less sensitive and these schools should become the investment targets in the investment decision-making. Therefore, it further explained that the comprehensive evaluation model was more reliable.

Reference
[6]. Information on http://www.nces.ed.gov/ipeds
[7]. Information on http://collegescorecard.ed.gov