International Journal of Advanced and Applied Sciences, 3(4) 2016, Pages: 1-6

Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Economic evaluation of world cities based on factor analysis

Anita Bai *, Swati Hira, P.S. Deshpande

Department of Computer Science & Engineering, Visvesvaraya National Institute of Technology, Nagpur, India

ARTICLE INFO

Article history:
Received 20 February 2016
Received in revised form
5 May 2016
Accepted 7 May 2016

Keywords:
Factor analysis
Economic parameters
City economy
Comprehensive analysis
Economic development

ABSTRACT

The economic performance of a city is based on various parameters such as industrial, transport, financial, and health etc. It also helps to evaluate economic and financial growth of a city. This paper describes factor analysis application to perform a comprehensive analysis and to evaluate the economic development of 18 cities on 15 different parameters. The performance of our approach is evaluated using IMF dataset for 18 world cities and the result shows the economic status of cities. We saw that our calculated rank and the rank provided by world ranking list are almost similar which proves that the analysis of world cities economic evaluation is successfully based on factor analysis.

1. Introduction

With the development of the economy, the competition between cities is increasing. We analyze the competitiveness between cities based on economic growth for a comprehensive, objective and systematic evaluation. Economic parameters play an important role in economic growth and it also affects economic performance of a country (Bai et al., 2015) which can be evaluated using various techniques such as ANOVA (Hair et al., 2006), Multidimensional modeling (Hira and Deshpande, 2012) etc. Every economic parameter interprets some valuable information from different aspects of evaluation, but there are certain relationships exist between parameters. This information will generate overlapping results during analysis process (Yuxiulin, 2006). Factor analysis method can be used to overcome these problems efficiently and provide level wise evaluation of economic development between cities. Economic development helps to improve the living standard of a city. A nation's standard of living is determined by stability in the economic productivity that meets the test of international markets.

The main objective of our paper is to evaluate the level of economic competitiveness and to rank cities on the basis of the comprehensive analysis. It is reasonable to expect that this analysis supports private-public investment, economic-and political decisions, entrepreneurship and social progress etc.

2. Preliminaries: Factor analysis

Factor analysis (Hair et al., 2006; Rencher, 2003) is used to uncover the latent structure (dimensions) of a set of variables. It reduces attribute space from a larger number of variables to a smaller number of factors. Factor analysis has a variety of applications such as an assessment of underlying relationships or dimensions in the data, and their placement of original variables with fewer, new variables.

The factor analysis model (Rencher, 2003; Wu and Zhang, 2003) expresses each variable as a linear combination of underlying *common factors* $f1, f2, ..., fm$, with an accompanying error term to account for that part of the variable that is unique (not in common with the other variables). For $y1, y2, ..., yp$ in any observation vector **y**, the model is as follows:

$$\begin{cases} y1 - \mu 1 = \lambda 11\, f1 + \lambda 12\, f2 + \cdots + \lambda 1m\, fm + \varepsilon 1 \\ y2 - \mu 2 = \lambda 21\, f1 + \lambda 22\, f2 + \cdots + \lambda 2m\, fm + \varepsilon 2 \end{cases}$$
$$yp - \mu p = \lambda\, p1\, f1 + \lambda\, p2\, f2 + \cdots + \lambda\, pm\, fm + \varepsilon\, p \qquad (1)$$

Ideally, m should be substantially smaller than p; otherwise we have not achieved a parsimonious description of the variables as functions of a few underlying factors. We might regard the fs as random variables that engender the y's. The coefficients λij are called *loadings* and serve as weights, showing how each yi individually depends on the fs.

3. Data normalization and preprocessing

3.1. Data

Yearly data on 15 economic parameters were obtained from the IMF Finance data (IMF, 1945) over the period 2008–2012 for the 18 cities (Hong Kong, New York, London, Singapore, San Francisco,

* Corresponding Author.
Email Address: anita.bai@students.vnit.ac.in (A. Bai)

Sydney, Toronto, Los Angeles, Chicago, Shanghai, Tokyo, Seoul, Boston, Geneva, Dubai, Berlin, Moscow and Zurich). The data have been transformed to normalized form. Table 1, describes the IMF sample dataset use for further processing.

Table 1: Sample dataset

City/Series-specific Notes	2008	2009	2010	2011	2012
GDP per capita	3,70,348.38	3,28,165.04	3,87,194.03	3,61,730.51	3,76,429.83
Gross national savings (% of GDP)	45.238	68.002	49.231	60.359	61.936
Volume of Imports of goods and services (% change)	43.363	7.972	-14.844	88.078	17.292
Retail sales (% change)	64.06	6.455	44.216	18.395	75.755
Consumer price index (% change)	92.167	91.469	29.672	29.749	28.893
Industrial production (% change)	34.762	52.198	28.938	98.499	20.883

3.2. Data normalization

Data normalization is used for proper organization of data. We use $Z_{min-max}$ normalization method, which transforms the data to scale within the range of [0, 1].

$$X_{norm} = (X_{mn} - min)/(max - min) \quad (2)$$

Where, X_{norm}, result of the normalized value of parameter,
X_{mn}, parameter data value to be normalized,
max, upper bound of the parameter data value,
min, lower bound of the parameter data value.

4. Evaluation of ranking of world cities

This paper applies factor analysis on selected 18 cities data to perform a comprehensive evaluation of their economic growth. We use IBM SPSS software to calculate and analyze the correlation coefficient matrix of 15 parameters and these parameters having a strong correlation among them. Therefore, Factor analysis can be used to minimize and classify the total number of parameters.

4.1. Notations

This section describes the economic parameters used to evaluate the economic status of world cities. Here V_1, V_2, ..., V_{15} describes the abbreviations used for parameters.

- V_1- Gross domestic product per capita current prices (National currency);
- V_2- Interest rate (National currency per current international dollar);
- V_3- Consumer price index (Percent change);
- V_4- Gross national savings (Percent of GDP);
- V_5- Volume of Imports of goods and services (Percent change)
- V_6- Volume of exports of goods and services (Percent change);
- V_7- Retail sales (Percent change);
- V_8- Education (Persons);
- V_9- Population (Persons);
- V_{10} - Transportation (National currency);
- V_{11} - General government total expenditure (National currency);
- V_{12} - General government revenue (National currency);
- V_{13} - General government structural balance (National currency);
- V_{14} - Tax charges (National currency);
- V_{15}- Industrial production (Percent change);

Mostly economic development of cities is described in terms of above selected parameters. GDP is one of the primary indicators of a country's economic performance and used to measure the total output of a country. General government total expenditure is generally categorized into expenditures on administration, defense, internal security, population, unemployment, health, education, foreign affairs, etc. General government revenue is used to finance the goods and services delivered to citizens and businesses through government and to fulfill their redistribution. It also provides an indication of the importance of the public sector in the economy in terms of available financial resources. The government should focus on other economic parameters such as interest rates, gross national savings, retail sales, education, general government structural balance, tax charges etc. to determine economic status of cities.

4.2. Evaluation of correlation coefficient matrix

Data validity and reliability is a preliminary step before performing factor analysis (Lulu, 2010). Reliability indicates the consistent degree of measurement. We measure the reliability of the data using Cronbach coefficient α. Validity indicates the closeness of the measured values. We calculated validity using KMO and Bartlett's test of Sphericity. KMO statistic is mainly used to test the simple and partial correlation between parameters, when the value of KMO is simple, indicating that these parameters are not suitable for factor analysis. Bartlett's test of Sphericity measures whether the parameters of the correlation matrix are independent or not. If the value of Bartlett's test of Sphericity is large, then reject the original hypothesis. It represents the parameters are related and can be used for factor analysis.

We used IBM SPSS software to analyze reliability and validity of data. Results show that α coefficient value is 0.807 which presents data reliability as good. KMO statistic value is 0.712; indicating factor analysis is relatively suitable. The significance probability of Bartlett's test of Sphericity is $0.00 < 0.01$, rejects the initial hypothesis which shows that the parameters are related to each other, so factor analysis is suitable to apply. Table 2 represents the test results.

4.3. Evaluation of correlation matrix

According to the correlation matrix shown in Table 3, it is analyzed that most of the parameters are strongly correlated, so it is essential to make use of factor analysis.

4.4. Evaluation of Eigen value λi, correlation matrix for contribution rate of variance, and extracted factors

Table 4 represents Eigen value λi and correlation matrix for contribution rate of variance. We extract common factors using principal component analysis method. From Table 4, there are three common factors and their cumulative variance proportion has reached 89.206%. So we represent the whole amount of information using three selected common factors.

Table 2: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.712
Bartlett's Test of Sphericity	Chi-Square (Approx.)	336.404
	df	74
	Sig.	.000

Table 3: Correlation matrix

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}
V_1	1	0.59	0.29	0.77	0.81	0.23	0.53	0.53	0.17	0.58	0.75	0.51	0.37	0.42	0.53
V_2	0.59	1	0.73	0.63	0.67	0.52	0.62	0.58	0.35	0.65	0.64	0.87	0.41	0.45	0.52
V_3	0.29	0.73	1	0.88	0.45	0.79	0.57	0.52	0.27	0.59	0.53	0.43	0.98	0.46	0.52
V_4	0.77	0.63	0.88	1	0.76	0.73	0.99	0.96	0.68	0.99	0.89	0.78	0.69	0.79	0.90
V_5	0.81	0.67	0.45	0.76	1	0.85	0.71	0.75	0.35	0.79	0.93	0.70	0.53	0.61	0.76
V_6	0.23	0.52	0.79	0.73	0.85	1	0.72	0.70	0.28	0.75	0.79	0.66	0.82	0.54	0.65
V_7	0.53	0.62	0.57	0.99	0.71	0.72	1	0.95	0.67	0.97	0.83	0.77	0.68	0.77	0.96
V_8	0.53	0.58	0.52	0.96	0.75	0.70	0.95	1	0.65	0.94	0.87	0.73	0.62	0.74	0.85
V_9	0.17	0.35	0.27	0.68	0.35	0.28	0.67	0.65	1	0.63	0.53	0.41	0.36	0.87	0.56
V_{10}	0.58	0.66	0.59	0.99	0.79	0.75	0.97	0.94	0.63	1	0.91	0.80	0.70	0.80	0.93
V_{11}	0.75	0.64	0.53	0.89	0.93	0.79	0.83	0.87	0.53	0.91	1	0.73	0.63	0.74	0.90
V_{12}	0.51	0.87	0.43	0.78	0.70	0.66	0.77	0.73	0.41	0.80	0.73	1	0.52	0.59	0.67
V_{13}	0.37	0.41	0.98	0.69	0.53	0.82	0.68	0.62	0.36	0.70	0.63	0.52	1	0.52	0.62
V_{14}	0.42	0.45	0.46	0.79	0.61	0.54	0.77	0.74	0.87	0.80	0.74	0.59	0.52	1	0.73
V_{15}	0.53	0.52	0.52	0.90	0.76	0.65	0.96	0.85	0.56	0.93	0.90	0.67	0.62	0.73	1

Table 4: Total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.589	69.956	69.956	10.589	69.956	69.956	5.281	34.640	34.640
2	1.624	9.592	78.988	1.624	9.592	78.988	4.745	32.436	66.786
3	1.325	8.678	88.286	1.325	8.678	88.286	3.212	22.410	89.206
4	.810	5.335	95.531						
5	.353	2.985	96.405						
6	.327	1.516	97.921						
7	.155	1.164	98.026						
8	.095	.468	99.653						
9	.050	.270	99.913						
10	.008	.050	99.973						
11	.002	.012	99.994						
12	.001	.006	99.999						
13	7.677E-05	.003	100.000						
14	1.341E-05	9.609E-05	100.000						
15	-4.545E-16	-3.030E-15	100.000						

Table 5 represents the communality of 15 parameters. The initial column values of 15 parameters from principal component analysis represent the communalities of the original parameters are 1. Extraction column values show the parameter communalities, calculated by three common factors. So we can confirm that the extracted common factors perform a good analysis of the 15 parameters.

4.5. Calculation of the common factors as rotated component matrix

To explain the common factors, primary structure of Component Matrix should be understandable which requires the rotation of Component Matrix.

The objective of the rotation is to place the axes close to as many points as possible. If there are clusters of points (corresponding to groupings of *y*'s), we seek to move the axes in order to pass through or near these clusters. This would associate each group of parameters with a factor (axis) and make interpretation more objective. The resulting axes then represent the natural factors. That is, in some common factors all parameters have high loads, while in others smaller loads. We use Varimax with Kaiser Normalization method to obtain Rotated Component Matrix is shown in Table 6.

Table 5: Parameters common level

Communalities		
	Initial	Extraction
Gross domestic product per capita current prices	1.000	.689
Interest rate	1.000	.737
Consumer price index	1.000	.964
Gross national savings	1.000	.957
Volume of Imports of goods and services	1.000	.877
Volume of exports of goods and services	1.000	.920
Retail sales	1.000	.926
Education	1.000	.898
Population	1.000	.902
General government revenue	1.000	.973
General government total expenditure	1.000	.924
Transportation	1.000	.770
General government structural balance	1.000	.987
Tax charges	1.000	.834
Industrial production	1.000	.824

Table 6: Factor loading matrix after rotation

Rotated Component Matrix			
	Component		
	1	2	3
Gross domestic product per capita current prices	.827	.108	.145
Interest rate	.819	.224	.140
Consumer price index	.261	.198	.957
Gross national savings	.559	.727	.373
Volume of Imports of goods and services	.775	.299	.292
Volume of exports of goods and services	.683	.204	.631
Retail sales	.517	.715	.368
Education	.523	.710	.327
Population	.101	.927	.068
General government revenue	.602	.691	.367
General government total expenditure	.728	.531	.334
Transportation	.761	.412	.228
General government structural balance	.237	.308	.914
Tax charges	.320	.875	.203
Industrial production	.528	.655	.340

The Rotated Component Matrix can be analyzed from the following aspects: Taking the loading of the factors into consideration, the first key factor is mainly related to Gross domestic product per capita current prices, Interest rate, Volume of Imports of goods and services, Volume of exports of goods and services, General government total expenditure and Transportation.

Second key factor is primarily related to Gross national savings, Retail Sales, Education, Population, General government revenue, Tax charges and Industrial production. Third key factor is related to Consumer price index and General Government structural balance. Here only the first three key factors were considered, because the total variance of the original parameters can be explained about 89.206% by the first three key factors.

4.6. Covariance matrix of factor score

Table 7 indicates that the 3 common factors extracted are not related to each other, and represents a score covariance matrix in the form of unit matrix.

4.7. Visual results of factor analysis

Fig. 1 shows the Eigen value and the number of factors in the Y-axis and X-axis respectively. So we can observe that the Eigen values of the top 3 factors are higher than 1, the polyline is steep, and the line tends to be stable from the fourth factor. Thus 3 factors should be selected for visual observation.

Table 7: Component score covariance matrix

Component	1	2	3
1	1.000	.000	.000
2	.000	1.000	0.000
3	.000	0.000	1.000

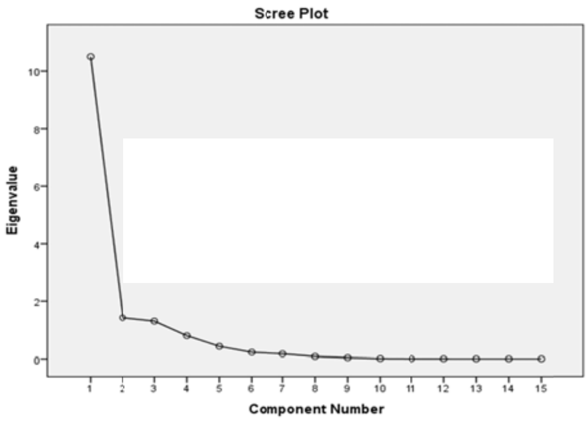

Fig. 1: Determining number of factors

4.8. Calculation of the component score coefficient matrix and comprehensive scores

Table 8 represents the calculation of component score coefficient matrix for 18 cities. Every factor score of 18 cities can be calculated according to Table 8:

$$F_1 = .333V_1 + 0.302V_2 - 0.173V_3 + \cdots \ldots - 0.107V_{14} - 0.019V_{15}, \quad (3)$$

$$F_2 = -.182V_1 - .118V_2 - 0.112V_3 + \cdots \ldots + 0.294V_{14} + 0.128V_{15}, \quad (4)$$

$$F_3 = -0.120V_1 - 0.124V_2 + 0.514V_3 + \cdots \ldots - 0.065V_{14} - 0.009V_{15}, \quad (5)$$

V_1, V_2, V_3, ..., V_{14} and V_{15} respectively indicate Gross domestic product per capita, Interest rate, Consumer price index, ..., Tax charges and Industrial production in the standardized values of original parameters.

In order to evaluate the economy rank of 18 cities, we calculate their comprehensive scores. The outcome of Table 9 is calculated using given formula.

Comprehensive evaluation score of a city:

$$CS_i = \frac{\sum_1^n F_i}{\sum_{j=1}^m \sum_{i=1}^n SF_{i,j}} + CAB_i, \quad (6)$$

Where,
m=No of cities
n=No of factors
$SF_{i,j}$ = Sum of all factor values in 18 cities.
CAB_i= Current account balance of a city (percent change).

Table 8: Component score coefficient matrix

	Component		
	1	2	3
Gross domestic product per capita current prices	.333	-.182	-.120
Interest rate	.302	-.118	-.124
Consumer price index	-.173	-.112	.514
Gross national savings	.002	.149	-.004
Volume of Imports of goods and services	.261	-.105	-.038
Volume of exports of goods and services	.102	-.171	.262
Retail sales	-.011	.163	.017
Education	.040	.153	-.026
Population	-.241	.434	-.116
General government revenue	.038	.120	.002
General government total expenditure	.144	.029	-.033
Transportation	.228	-.030	-.096
General government structural balance	-.153	-.067	.483
Tax charges	-.107	.294	-.065
Industrial production	.019	.128	-.009

Table 9: Common factors' scores and comprehensive scores

Cities	F1	F2	F3	Comprehensive Scores	Rank
New York	0.013962	0.018763	-0.00265	3.392984	1
London	0.018905	0.019563	-0.00569	1.956189	2
Hong Kong	-0.01389	0.07679	-0.03278	1.353501	3
Zurich	0.01489	0.036309	0.005823	1.179426	4
Singapore	0.363264	-0.18251	-0.10895	0.923193	5
Tokyo	-0.06969	0.196954	-0.04243	0.790956	6
Shanghai	-0.23247	0.51075	-0.11934	0.379406	7
Seoul	-0.03163	0.243119	-0.03537	0.376509	8
Geneva	0.468994	-0.08934	-0.13972	0.290763	9
San Francisco	0.601443	-0.21333	0.11796	0.147256	10

Table 9, represents the comprehensive scores of top 10 economic developed cities; however, we calculated the results for 18 cities. San Francisco had the lowest position among all city economies while the New York economy is high, because of its advanced infrastructure, education and technology.

5. Conclusion

Factor Analysis is a statistical technique and we used it to calculate common variances in the evaluation of city economic development. First, we classify and interpret parameters in the economic evaluation by factor analysis. Then we extract a number of common factors, and finally calculate rank according to their scores evaluated by comprehensive score model. From our analysis, we observed that our calculated rank list (New York, London, Hong Kong, Zurich, Singapore, Tokyo, Shanghai, Seoul, Geneva and San Francisco) and the rank provided by world ranking list (New York, London, Hong Kong, Singapore, Zurich, Tokyo, Seoul, Boston, Geneva and San Francisco) of top richest cities are almost similar. So from our analysis, we conclude that it is successful to apply factor analysis to evaluate the estimated rank of economically richest cities. Since the growth of an industry in any city depends upon its economy, so our approach will also help experts to analyze their growth in a city or to set up an industry and explore their business. Extracted factors make the country attractive for foreign direct investments, employment, higher education and technology.

Acknowledgement

The authors would like to thank faculties of Department of Computer Science and Engineering, VNIT, Nagpur, for their valuable guidance and support for this work.

References

Bai A, Hira S, Gowardhan R and Deshpande PS (2015). The impact of economic parameters on government total expenditure in 5 richest coun tries. Australian Journal of Basic and Applied Sciences, **9**(6): 77-82.

Hair JF, Black WC, Babin BJ, Anderson RE and Tatham RL (2006). Multivariate data analysis. 6th Edition, Upper Saddle River, NJ: Pearson Prentice Hall, New York, USA.

Hira S and Deshpande P S (2012). Analysis of various economic parameters using multidimensional analysis and data mining. IDRBT Doctoral Colloquium, Hyderabad, India.

IMF dataset (1945). Available: http: //www.imf.org/ external/data.htm.

Lulu Q (2010). Research on satisfaction influencing factors of livable city based on structural equation modeling. M.Sc. Thesis, Chongqing University, Chongqing, China.

Rencher AC (2003). Methods of multivariate analysis. John Wiley & Sons, New York, USA.

Wu N and Zhang J (2003). Factor analysis based anomaly detection. In Information Assurance Workshop. IEEE Systems, Man and Cybernetics Society, 108-115.

Yuxiulin R (2006). Multiple Statistical Analysis. Beijing: China Statistics Press, Beijing, China.