International Journal of Industrial Engineering and Management (IJIEM), Vol. 2 No 2, 2011, pp. 45-50
Available online at http://www.ftn.uns.ac.rs/ijiem/
ISSN 2217-2661

# Using Patent Statistics and Principal Component Analysis to Predict Global Competition

**François A. Ravalison**
Ecole Supérieure Polytechnique, University of Antananarivo, Madagascar, fravalison@moov.mg

**Narisoa Rabenja**
Malagasy Intellectual Property Office-Antananarivo, Madagascar

Received (14 February 2011); Revised (13 Jun 2011); Accepted (29 Jun 2011)

**Abstract**

*Globalization and acute competition require continuously searching new approach in identifying ways for better economic growth. The purpose of this paper is to identify a process to predict global competition. The data are collected from the Malagasy Intellectual Property Office. Frequencies of foreign patents, registered at that Office, from 1994 to 2009 have been collected. Then, data mining is conducted to bring out an idea. Findings reveal that Competition indicators based on patent statistics are confirmed as appropriate measure of competition. Then Principal Component Analysis (PCA), applied on Competition indicators, permits to predict sector of global competition in a country level. This paper serves as a valuable analytical framework for the management of patent data for continuous innovation for economic growth.*

**Key words:** *Competition Indicator, Global Competition, Patent, Principal Component Analysis*

## 1. INTRODUCTION

Global competition begins to be present on the developing countries' market. It affects all sectors of economy. Africa in general and Madagascar in particular cannot avoid that economical trend. Lack of appropriate strategy has thrown some national or private enterprises in difficulties. Others have closed their doors. And then, unemployment rate increases.
In fact, enterprises and policy makers are not well informed. There is no reliable indicator which permits to measure and to predict global competition. Sometimes, data exist but they are not sufficiently analyzed and outputs are not complete and nor predictive. Even some foreign enterprises, for example French banks in Madagascar, utilize competitive intelligence but they do not have reliable Critical Success Factors to gauge and to anticipate even local competition. Foreign industries vulnerable points are global and local competition because they are not well informed in advance.
Two research questions appear: How can global competition be measured? How to determine the global competition trends?
The objectives of this study are:

- to identify measurement of global competition,
- to identify direction of global competition.

## 2. LITTERATURE REVIEW

### 2.1 *Patent concerns knowledge creation*

Patents are an objective measure of R&D activities [2]. If that the case, we call it "patent activity". The ability of patent information to measure R&D activities is attributed to some characteristics of patents [2]. A large amount of technological information is contained in patents [3] since R&D is concerned. For that, they are often the only source for the timely recognition of technological change [1] [2].

### 2.2 *Patent concerns performance and competition*

R&D indicates technological activities which lead to subsequent market [3]. And when a firm changes its R&D expenditures, parallel changes occur in its level of patenting [4]. Patent quality measures the impact of those activities [3]. Besides, a patent present a patentee's non-negligible expectation as to the ultimate utility and marketability of the invention [3]. And Ernst [2] has establishes a positive relationship between patent applications and subsequent sales increase. And then, he [3] has established empirical evidence and suggested a positive relationship between patenting strategies and company performance.

### 2.3 *Previous research about patent and competition-a Competition Indicator*

Banerjee et al. have conducted significant research about patent and competition. They say that a patent is an input to the processes of production or services. And patents statistics permit understanding the degree of competition and the competition-driven research

strategy [1]. Besides, those patent statistics can be used to understand the nature of competition through game theory modeling of patent participants and the competition in which they are involved [1]. Many scholars do in fact treat these statistics as indicator of “competitiveness” [8], [9], [12]
Banerjee et al. have proposed a set of indicators based on patent statistics [1]. Those indicators make known scale of competition between foreigner and national researchers. We suggest utilizing the following [10]:
Let ($t_{ij}$) represent the number of patents of type i in year j. We calculate:

$$T_j = \sum_i t_{ij} \quad (1)$$

as the total numbers of patents during year j.

And:

$$T^i = \sum_j t_{ij} \quad (2)$$

the total numbers of patents of type i over all years.
So:

$$T = T_j + T^i \quad (3)$$

as total of all types of patents over all years.
The competition indicator $I_{ij}$ for each patent of type i for each year j [10]:

$$I_{ij} = \frac{\frac{t_{ij}}{T_j}}{\frac{T^i}{T}} \quad (4)$$

The numerator shows the share of patent of type i for a year j. And the denominator shows the share of patent of type i of all types over all years.

### 2.4 Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables. Its goal is to extract the important information from the table, to represent it as set of new orthogonal variables called principal components. (PCA) also represents the pattern of similarity of the observations and the variables by displaying them as points in maps [6] [11].
Let use the following notations:

- **A** : matrix
- **a** : vectors
- x : elements of matrix
- **I** : identity matrix
- superscript $^T$: transpose operation

There are *I* observations and *J* variables in the data table represented by *IxJ* matrix X whose element is $x_{i,j}$. and the level or the rank of matrix X is *L* which is computed as:

*rank of matrix X = min {I,J} = L* (5)

In general, the data table will be processed before the analysis. The columns of X will be centered so that the mean of each column is equal to zero. It has the following singular decomposition [14]:

$$X = P\Delta Q^T \quad (6)$$

P: matrix *I*x*L* of left singular vectors.
Q: matrix *J*x*L* of right singular vectors.
Δ: diagonal matrix of singular values

$$\Delta^2 = \Lambda \quad (7)$$

Λ: diagonal matrix of eigenvalues of $X^TX$ and $XX^T$.
The inertia of a column is:

$$\gamma_j^2 = \sum_i^I x_{ij}^2 \quad (8)$$

And the inertia  of the data table or the total inertia is computed as:

$$J = \sum \gamma_{i,j}^2 \quad (9)$$

The *Center of Gravity g* of the rows is the vector of the means of each column of X. and the Euclidean distance of the *i*-th observation to *g* is equal to:

$$d_{ij}^2 = \sum_j^J (x_{i,j} - g_j)^2 \quad (10)$$

In PCA, the components are obtained from the singular decomposition of the data table X. From Takana decomposition above, the *I*x*L* matrix of *factor scores* is calculated by:

$$F = P\Delta \quad (11)$$

The matrix Q gives the coefficients of the linear combinations used to compute the *factors scores*:

$$F = P\Delta = P\Delta QQ^T = XQ \quad (12)$$

Matrix Q is a projection matrix which transforms the original data into factor scores. This matrix can also be used to compute factor scores for observations. The latter are called *supplementary (sup) or illustrative observations*.
The component can also be represented geometrically by the rotation of the original axes.

with: $$F^TF = \Delta^2 \quad (13)$$

$$X = FQ^T \quad (14)$$

$$Q^TQ = I \quad (15)$$

This decomposition is called *bilinear decomposition* of X [7].

The observations to compute PCA are called *active observations.* The factor scores for supplementary observations are obtained by first positioning these observations into PCA space and then projecting them onto the principal components:

$$f_{\text{sup}}^{T} = X_{\text{sup}}^{T} Q \tag{16}$$

The goals of PCA are: to extract the most important information from the data table, to compress the size of the data set by keeping only this important information, to simplify the description of the data set, and to analyze the structure of the observations and the variables.

## 3. METHODOLOGY

### 3.1 *Research process*

The research process contained three principal steps. It is presented in the Figure 1 below.

There are three activities: collect frequency of patents deposited by Non Residents in Madagascar, calculate matrix of competition indicators and process the Principal Component Analysis.

The inputs of the first step are data from Malagasy Intellectual Property Office. Data, concerning patent deposited by Non Residents (NR) during 1994-2009, have been captured in Excel. A table with three columns such as Year, International Patent Classification (IPC), Country of Origin, has been obtained. So per year, frequency of patents deposited is obtained. The output of that first step is Patent Statistics.

In the present case, Patent Statistics is a two dimensional array. On the basis of those statistics, Competition Indicators [1] are calculated at the second step. Banerjee calculation process is used [1]. The output of the second step is a matrix of Competition Indicators.

The Principal Component Analysis of that matrix is processed at the third step.

### 3.2 *Software tool*

Data have been captured in Excel 2007. And XLStat 6.0 has been utilized to process the Principal Component Analysis.

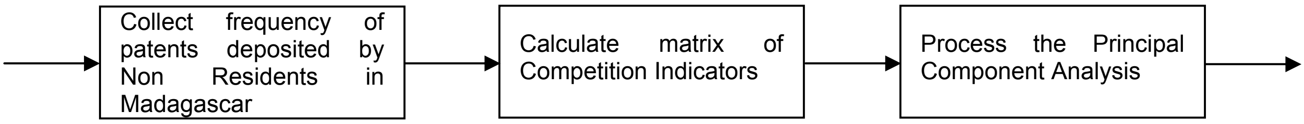


**Figure 1.** Research Process

## 4. FINDINGS

The output of the second step of the research process is a matrix of Competition Indicators: eight columns of IPC and sixteen rows of years. It has been charted and has permitted to obtain the following Figure 2 of trend over time of Competition Indicators.

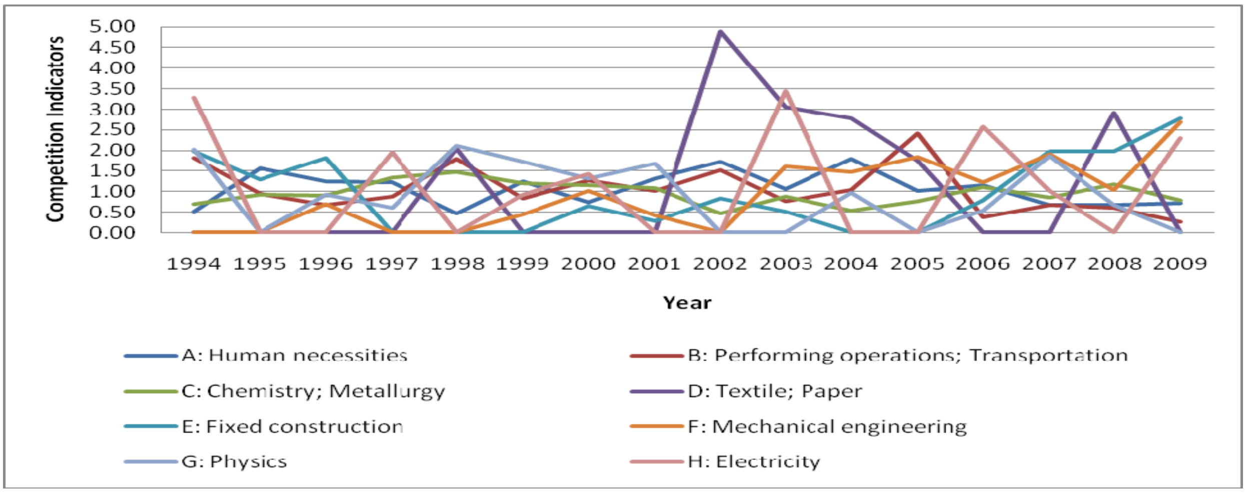


**Figure 2.** Trend Over Time of Competition Indicators

There are eight lines corresponding to eight patent sectors: Human necessities, Performing operations and Transportation, Chemistry and Metallurgy, Textile and Paper, Fixed construction, Mechanical engineering, Physics, Electricity. The majority of CI is superior to 0 and inferior than 2. Five sectors have their CI>2: Electricity, Textile and Paper, Performing operations and Transportation, Fixed construction, and Mechanical engineering. Textile and Paper is a cut above the rest. Its CI was at 5 in 2002. A first cluster of sectors has trendline with gentle slope<0: Human necessities, Performing operations and Transportation, Chemistry and Metallurgy, and Physics. A second cluster has trendline with gentle slope>0: Fixed construction, and

Electricity. And a third one has trendline with steep slope>0: Textile and Paper, Mechanical engineering. We notice that the third cluster is competitive followed by the second cluster. The first one is not competitive.

At the third step, we have processed with covariance as type of matrix and variances with 1/n. The number of factors associated with non-trivial eigenvalues is 7, and 1 is removed. The following Fig. 3 presents the Principal Component Analysis of Competition Indicators.

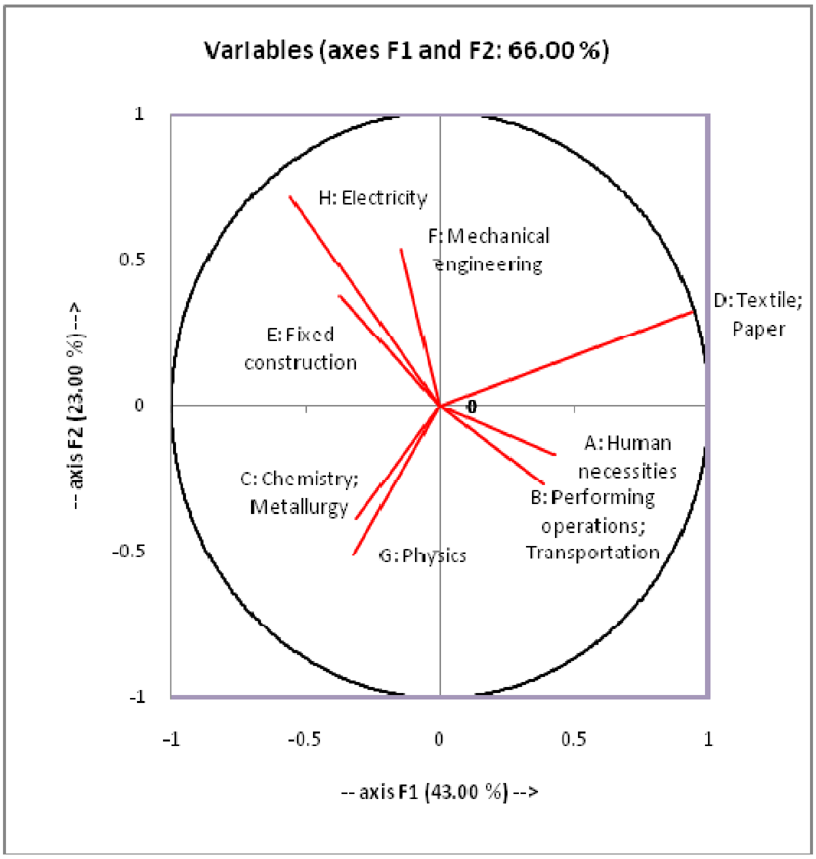


**Figure 3.** Principal Component Analysis of Competition Indicators

There are four clusters of IPC in each quadrant. In the first quadrant, there is only one IPC relating to Textile and Paper. In the second, there are three IPCs: Electricity, Mechanical engineering and Fixed construction. There are two IPCs in the third and fourth quadrants: Chemistry-Metallurgy and Physics for the first one and Human necessities and Operations-Transportation for the second one.

## 5. DISCUSSION

Global competition can be measured by Banerjee's Competition Indicators. And Principal Component Analysis permits to determine the cluster of global competition trends.

Banerjee survey concerns the sector of biotechnology because it is an emerging area with acute competition [1]. Besides, it is a very high expectation on the future economic returns [1]. In our case, we have considered all sectors define by International Patent Classification. Our area of global competition is focused in the case of Madagascar. That global competition is measured by Competition Indicators. Different values for different IPC are obtained. Trend over time of those indicators are too difficult to read and to understand. The corresponding result is acceptable but confused. Because the values of indicators are alike. Maping of Banerjee's CI shows a certain limit in terms of Global Competition measurement. Although trend over time shows the variations of each variable.

To better understand the competition situation, we have determined the global competition trends. For that, we have used Principal Component Analysis. The corresponding results are acceptable and clear. They confirm that Banerjee's theory on Competition Indicators can measure global competition at a local level. Besides, application of Principal Component Analysis permits to map those trends and to present some clusters. That application contributes to some explanation about some problematic relating to Competition Indicators [10]. It is said that the above theory, we mean Banerjee's, does not show real competition. In our case and precisely with our trend over time, the theory permits to appreciate and to understand competition through a global view when mapping is performed. But it is not sufficient to assess competition.

Our research shows that Competition Indicators can be processed to predict competition trends. The Principal Component Analysis shows that it is an appropriate tool to map factors' profile [13] and to determine area of competition. In our case, it shows that "Textile; Paper" is a domain with acute competition.

## 6. CONCLUSION

In this paper, Competition Indicators have been used to map competition through patent statistics in an African developing country. The obtained mapping is confused, in terms of variations, because of indicators' values. Understanding has been improved. And the Principal Component Analysis has been used to improve the mapping and to get a new and clear idea concerning the competition trends.

Practical implications:

- The research output can be used by policy makers to guide their strategy. For example: in what domain a country will focus its competitive advantage? Or in what sector should Government bring a certain attention?
- The research output is a practical tool to shape the future in terms of competition.

Originality/value:

- Contribution of an "added value", through a statistic tool, to a past research.
- Creation of an engineering process, matrix of Banerjee's Competition Indicators process and PCA process, to predict competition.

Future research:

- Competition is more or less determined, but how will we determine the speed of that competition?
- And how will we measure national vs. international competition forces, based on patent statistics, when Banerjee's Competition Indicators are known?

## 6. ACKNOWLEDGEMENTS

We would like to thank the Malagasy Intellectual Property Office. Its Director General has authorized us to access their precious data. He has accepted one of his staff to collaborate with us to perform the present paper.

## 7. REFERENCES

[1] Banerjee P., Gupta B. M., Garg K.C. (2000), "*Patent Statistics as Indicators of Competition: an Analysis of Patenting in Biotechnology*", Scientometrics, Vol. 47, pp. 95-116.

[2] Campbell, R.S. (1983), "*Patent Trends as a Technological Forecasting Tool*", World Patent Inf., N°5, Vol.3, pp.137–143.

[3] Ernst, H. (1996), "*Patentinformationen für die Strategische Planung von Forschung und Entwicklung*", DUVVerlag-Wiesbaden.

[4] Ernst, H. (1998), "*Patent Portfolios for Strategic R&D Planning*", Journal of Engineering and Technology Management, Vol.15, pp. 279-308.

[5] Griliches, Z., Pakes A., Hall B.H. (1986), "*The Value of Patents as Indicators of Inventive Activity*", Discussion Paper No. 1285, Harvard Institute of Economic Research, Cambridge, MA.

[6] Jolliffe I.T. (2002), "*Principal Component Analysis*", New York, Springer.

[7] Kruskal J.B. (1978), "*Factor Analysis of Principal Component Analysis: Bilinear Methods*", International Encyclopedia of Statistics, pp. 307-330, New York, The Free Press.

[8] Pavitt K., Patel P. (1988), "*The International Distribution of Determinants of Technological Activities*", Oxford Review of Economic Policy, Vol. 4, N°4, pp. 35-55.

[9] Pavitt K. (1985), "*Patents Statistics as Indicators of Innovative Activities: Possibilities and Problems*", Scientometrics, Vol. 7, pp. 77-99.

[10] Ramani S. V., De Looze M. A. (2000), "*A Note on Using Patent Statistics to Obtain Competition Indicators*", Scientometrics, Vol. 49, N°3, pp. 511-515.

[11] Saporta G., Niang N. (2009), "*Principal Component Analysis: Application to Statistical Process Control*", Data Analysis,pp. 1-23.

[12] Soete L. (1987), "*The Impact of Technological Innovation International Trade Patterns: the Evidence Reconsidered*", Research Policy, Vol. 16, N° 2-4, pp. 10.

[13] Suhr D.D. (2007), "*Principal Component Analysis vs. Exploratory Factor Analysis*", Statistics and Data Analysis, Paper 203-30.

[14] Takana Y. (2002), "*Relationship Among Various Kinds of Eigenvalue and Singular Decompositions*", New Development Psychometrics, pp. 45-56.