Physica A 392 (2013) 177–183

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

An ideal gas approach to classify countries using financial indices

Paulo S.G. de Mattos Neto [a], George D.C. Cavalcanti [a], Francisco Madeiro [b], Tiago A.E. Ferreira [c,*]

[a] *Center of Informatics, Federal University of Pernambuco, Recife, Brazil*

[b] *Department of Computer Science, Catholic University of Pernambuco, Brazil*

[c] *Department of Statistics and Informatics, Federal Rural University of Pernambuco, Brazil*

ARTICLE INFO

Article history:
Received 26 March 2012
Received in revised form 10 July 2012
Available online 27 July 2012

Keywords:
Econophysics
Countries classification
Financial indices
Maxwell–Boltzmann distribution

ABSTRACT

Traditionally, countries' development is classified based on several features that can be related to economic and social factors. However, this classification task is costly due to the difficulty of obtaining those features. We propose a method to classify countries based on financial indices using an ideal gas model. The probability density function (pdf) of the return series of the financial indices is used to characterize the fluctuation of a market. Based on the pdf, the volatility and the *B* coefficient, which describe the behavior of world markets, are estimated. The evaluation procedure uses 34 indices from developed and developing countries. The results show that the proposed method is an accurate, fast and low-cost computational alternative to the classifications provided by traditional organizations.

1. Introduction

Several proposals for classification or clustering of countries [1] are published every year by institutions, such as the World Bank and United Nations. The objective of those publications is to measure the degree of development of the countries based on economic, demographics, health, cultural and educational indicators. A country classification report is important because it generates indices that are used to guide the adoption of public policies in order to improve the population's quality of life.

The proposals of countries classification found in the literature address several factors, such as economic [2–4] and social [5] aspects. Examples of economic indicators are per capita income, gross domestic product (GDP) and industrialization level [2,3]. The Human Development Index (HDI) [5] is an economic indicator that combines economic and other measures such as: life expectancy rates and educational indices, e.g., rates of schooling and literacy. In general, the acquisition procedure of those indicators is an expensive, difficult and often inaccurate task. As a consequence, the classification of countries is a challenging and complex process.

This work proposes a method to classify countries using non-conventional features, such as: volatility and *B* coefficient, which is the decay rate of the exponential function. These features describe the fluctuation behavior of the world market indices. The advantages of using these indicators instead of traditional ones are: (i) they are easy to calculate; (ii) since only two features are used, the computational requirements are low; (iii) the acquisition process of the features does not demand a high cost. The proposed approach analyzes the probability density function (pdf) of the return series [6–13] of

* Corresponding author. Tel.: +55 81 33206490; fax: +55 81 33206490.
E-mail addresses: psgmn@cin.ufpe.br (P.S.G. de Mattos Neto), gdcc@cin.ufpe.br (G.D.C. Cavalcanti), madeiro@dei.unicap.br (F. Madeiro), taef.first@gmail.com, tiago@deinfo.ufrpe.br, taef@deinfo.ufrpe.br (T.A.E. Ferreira).
URL: http://cin.ufpe.br/~viisar (G.D.C. Cavalcanti).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2012.07.049

world financial indices and models the dynamics of the market as an ideal gas [14]. In the proposed approach, each market is defined by two parameters (volatility and *B* coefficient). A clustering algorithm [15] constructs two groups: developed and developing countries. The nearest neighbor rule is used to classify a query country.

This paper is organized as follows. Section 2 describes the country classification measures provided by traditional organizations. Section 3 presents the proposed method. Sections 4 and 5 show the results and the final remarks, respectively.

2. Classifications according to international organizations

There are some disagreements when classifications released by institutions are compared. For example, the International Monetary Fund (IMF) identifies thirty four countries as advanced economies, while the World Bank (WB) identifies sixty six high-income countries. The criteria to define those lists are often subjective and controversial. In the next sections some classifications of traditional organizations are described.

2.1. Classification according to the International Monetary Fund (IMF)

The IMF publishes a World Economic Outlook [3] twice a year. The survey is conducted by the IMF staff and presents the global economic developments. The report gives an overview and a detailed analysis of the global economy. Basically, the issues discussed cover fields that affect the industrial countries and developing countries. The classification published in the World Economic Outlook divides the world into two major groups: advanced economies and developing countries (this group includes emerging economies and poor countries).

The clustering of countries is performed using several economic characteristics, such as exchange rates, interest rates, growth tax and GDP. Other data related with foreign trade and fiscal data are also used, as well as information about unemployment rate and employment rate. Thus, the groups defined by the IMF do not take into account social factors, such as education and life expectancy, for example.

2.2. Classification according to the World Bank (WB)

Generally, the World Bank's [2] main criterion for classifying the economies is the gross national income (GNI) per capita, which is also referred to as gross national product (GNP). Based on this information, every economy is classified as low income, middle income (subdivided into lower middle and upper middle), or high income.

The classification of the WB can also be divided into developing and developed economies. In this case, low-income and middle-income economies are referred to as developing economies. This nomenclature is interesting because other international organizations also use it, such as the UN and IMF. However, the World Bank highlights that not all economies in this group have the same level of development, because the classification by income does not necessarily reflect the development status of a country.

2.3. Classification according to the United Nations (UN)

The United Nations uses two types of classification: an economic classification, which is based on IMF, World Bank and other data, and a social classification, which is based on indices of education, wealth and quality of life in general. The economic classification is very similar to that of IMF, while the social classification estimated by UN is determined by the Human Development Index [5].

2.3.1. Human Development Index (HDI)

Traditionally, the classifications developed by international agencies highlight economic aspects. The Human Development Index of the United Nations is a statistical measure which estimates the level of human development of a particular country. This metric considers three dimensions: wealth, education and life expectancy. This index is a comparative measure and aims to estimate the welfare of a population.

The UN indicates that estimates of HDI have their focus in policies centered on people instead of GDP per capita, for example. The GDP measures the country performance by market value of all final goods and services produced in a country in a given period. Although there is some correlation between this index and a thriving economy, it does not necessarily follow that the richest countries have the highest HDI. However, in general, advanced countries classified as developed have a high Human Development Index.

3. Proposed method for the classification of countries based on the ideal gas theory

According to de Mattos Neto et al. [14], the company shares that are negotiated by investors can be compared with particles of an ideal gas model. Two assumptions that describe the dynamics of the markets can be used to justify this analogy. The first is the random walk hypothesis, in which there is no interaction between the particles. The second is the non-random walk hypothesis, in which there is a weak interaction between particles.

Following this assumption, the volatility can be associated with the temperature or thermal energy of the market. If the market is agitated, its temperature is high; otherwise, if the market is not agitated, its temperature is low. Therefore, from the point of view of an ideal gas, the market should be characterized by its temperature (or volatility).

Thus, the molecule velocity or energy of an ideal gas model can be compared with the stock prices. Just as a particle can change its energy (or speed), similarly, a stock can have its price changed. These changes in the value of the stock prices have the same behavior as a change in energy (or speed) of a particle. This relationship is observed when the pdf of the return series is analyzed. In the classical description, this pdf follows the Maxwell–Boltzmann distribution [16] that defines the probability function of the energy (or velocity) of each particle.

The Maxwell–Boltzmann distribution is defined by Eq. (1).

$$P(E) = a \cdot e^{-E/(k_b \cdot t)}, \tag{1}$$

where the coefficient a is a constant, k_b is the Boltzmann's constant [16], E is the energy of the system and t corresponds to the temperature.

In Ref. [14] the pdf of the return series was well fitted by an exponential function [17] that is described in Eq. (2).

$$P(x) = c \cdot e^{-B \cdot |x|}, \tag{2}$$

where the coefficients c and B are constants: c is the initial amplitude ($x = 0$) and B is the decay rate. The variable x corresponds to the return series, or system energy. With appropriate normalization, $c = B$ and the standard deviation of this exponential distribution is proportional to $1/B$ [18]. Therefore, B is inversely proportional to standard deviation or volatility.

By observing Eqs. (1) and (2), a correlation can be established. Therefore, t is proportional to the volatility of the system and E corresponds to values of the return series. The correlation between the B coefficient of the exponential function and the system temperature (Eq. (3)) can be extrapolated to the market system: when the volatility decreases (the thermal energy decreases or the temperature decreases), the B coefficient increases. This analogy can be associated with the system temperature.

$$B = \frac{1}{k_b \cdot t}. \tag{3}$$

Therefore, for a financial market

$$B \propto \frac{1}{Volatility} \Rightarrow B \cdot Volatility = C, \tag{4}$$

where C is the proportionality constant between the B coefficient and the volatility. Thus, the market system can be treated as a gas system.

Based on the temperature of the gas, or similarly, the temperature of the market, it is possible to cluster the economies according to the degree of agitation of each one. The elements of each cluster have common characteristics that can be used to analyze the markets. The idea is to use the analogy with the ideal gas to understand the behavior of markets through their fluctuation measures (B coefficient and volatility) obtained by the methodology herein proposed. This procedure gives information about the fluctuation/instability of markets using the pdf analysis of the return series of world indices.

The B coefficient and volatility is used to cluster economies. The degree of agitation of the indices is based on the temperature of the economies. In order to cluster the markets, the k-means clustering algorithm [15] can be used.

The proposed method is divided into two steps. First, the analogy previously described is used to analyze the pdf of the return series with the exponential function. From this analysis, two temperature measures are generated, volatility and B coefficient. The second step consists of clustering the world indices/stocks based on volatility and/or B coefficient.

In the first step, for each index the return series is defined as

$$G(t) = \frac{|\ln Z(t + \Delta t) - \ln Z(t)|}{\delta}, \tag{5}$$

where $\Delta t = 1$ day, $Z(t)$ is the series value at time t, and δ is the standard deviation of $|\ln Z(t + \Delta t) - \ln Z(t)|$. For each series, the volatility described in Table 1 is calculated as

$$V_t(t) = \frac{1}{n} \sum_{t=1}^{n} G(t), \tag{6}$$

where n is an integer that represents the number of time series observations. Then, the probability density function of the return series is estimated and the exponential function fitting is performed based on the Trust Region (TR) algorithm [19]. The Mean Square Error (MSE) measure is used to evaluate the fitting errors.

The fitting is performed using the TR algorithm, which presents a good performance as reported in Ref. [14]. The two variables (B coefficient and volatility), shown in Eqs. (3) and (6), represent the degree of the market agitation. As observed

Table 1
Fitting errors (MSE), financial features (volatility and *B* coefficient) and classification for country.

Indices (countries)	UN–IMF classifications	World Bank classification	HDI 2010	Volatilities of indices	Power law MSE	Exponential MSE	*B* coefficient
IGBC	Developing countries	Developing	0.689	1.1410	0.00021	0.00018	0.8945
IPSA		Developing	0.783	1.1290	0.00021	0.00010	0.8994
EZA		Developing	0.597	1.1270	0.00034	0.00023	0.9221
IGPA		Developing	0.783	1.1240	0.00029	0.00023	0.9050
CSI 300		Developing	0.663	1.1170	0.00032	0.00013	0.8876
SSEC		Developing	0.663	1.1041	0.00026	0.00006	0.9070
PSEI		Developing	0.519	1.0880	0.00039	0.00029	0.9958
Bse Sensex		Developing	0.519	1.0650	0.00114	0.00011	0.9448
JKSE		Developing	0.600	1.0600	0.00031	0.00029	0.9608
KWSE		Developed	0.771	1.0561	0.00522	0.00004	0.9336
SET		Developing	0.654	1.0380	0.00081	0.00013	0.9886
MXY		Developing	0.750	1.0230	0.00048	0.00033	1.0093
IPC		Developing	0.750	0.9772	0.00047	0.00034	1.0540
Merval		Developing	0.775	0.9597	0.00058	0.00017	1.0370
IBRX 50		Developing	0.699	0.9545	0.00030	0.00024	1.0550
MASI		Developing	0.567	0.9417	0.00013	0.00005	1.0780
Ibovespa		Developing	0.699	0.9258	0.00029	0.00004	1.0800
Hang Seng	Developed countries	Developed	0.862	1.0122	0.00029	0.00017	1.0260
OMX C20		Developed	0.866	1.0110	0.00030	0.00021	1.0324
MIB		Developed	0.854	1.0055	0.00121	0.00039	1.0240
IBEX 35		Developed	0.863	1.0054	0.00103	0.00013	1.0270
GSPTSE		Developed	0.888	0.9812	0.00003	0.00001	1.0340
ISEQ		Developed	0.895	0.9698	0.00308	0.00006	1.0585
CAC 40		Developed	0.872	0.9677	0.00160	0.00018	1.0690
Nikkei 225		Developed	0.884	0.9663	0.00024	0.00013	1.0580
SMI		Developed	0.874	0.9617	0.00020	0.00027	1.0520
FTSE 100		Developed	0.849	0.9599	0.00071	0.00022	1.0660
ATHEX 20		Developed	0.855	0.9592	0.00277	0.00004	1.1122
ASX 200		Developed	0.937	0.9559	0.00052	0.00009	1.0838
KOSPI		Developed	0.877	0.9367	0.00084	0.00014	1.0820
DJIA		Developed	0.902	0.9283	0.00061	0.00012	1.0570
S&P500		Developed	0.902	0.9181	0.00083	0.00018	1.0720
DAX 30		Developed	0.885	0.9130	0.00123	0.00031	1.1280
TWII		Developed	0.868	0.8935	0.00564	0.00005	1.0641

in Eq. (4), there is an inverse correlation between the *B* coefficient and the volatility and this information can be used to cluster the world markets.

In the second step of the method, feature vectors composed of two features, volatility and *B* coefficient, are used as input to the *k*-means algorithm. The *k*-means algorithm [15] consists of a method of cluster analysis that is applied to design centroids (or prototype vectors). Let $d(P, c_i)$ be the distance between P and c_i, where c_i is the i-th centroid and P is the data to be classified as member of the class i. The number of the centroids is defined by the user. Iteratively, the *k*-means adjusts the position of each centroid based on the data, minimizing the distance of the centroid c_i to the data of a class i. If $d(P, c_1) < d(P, c_2)$ then P is classified as a member of the class "1" (here, the class of developed countries). Otherwise, P is classified as a member of class "2" (here, the class of developing countries).

The proposed approach uses only two attributes to classify the countries. Another advantage is that the attributes generation is a simple process because only the financial characteristics of a given country are required. On the other hand, traditional classifications [2–4] use several social and economic indices. These indices are calculated based on a large amount of data, a fact that increases the cost. In addition, a high computational effort is needed to process this large number of indices.

4. Results

The return series of 34 world market indices in the period of two years (January 2008–January 2010) are used to evaluate the proposed approach. Half of the indices comes from developed markets and the other half from developing markets, based on the classification given by the UN and the IMF. The first group is composed of seventeen developed market indices: Hang Seng (Hong Kong), OMX C20 (Denmark), MIB (Italy), IBEX 35 (Spain), GSPTSE (Canada), ISEQ (Ireland), CAC 40 (France), Nikkei 225 (Japan), SMI (Switzerland), FTSE 100 (England), ATHEX 20 (Greece), ASX 200 (Australia), KOSPI (Korea), DJIA and S&P 500 (United States), DAX 30 (Germany) and TWII (Taiwan). The second group is composed of seventeen developing market indices: IGBC (Colombia), IGPA and IPSA (Chile), EZA (South Africa), CSI 300 and SSEC (China), PSEI (Philippines), Bse Sensex (India), JKSE (Indonesia), KWSE (Kuwait), SET (Thailand), MXY and IPC (Mexico), Merval (Argentina), IBRX 50 and Ibovespa (Brazil) and MASI (Morocco).

In our representation, each country is associated with one or more market indices. Table 1 shows the classification of each country (market index) given by the following international organizations: United Nations (UN), International Monetary

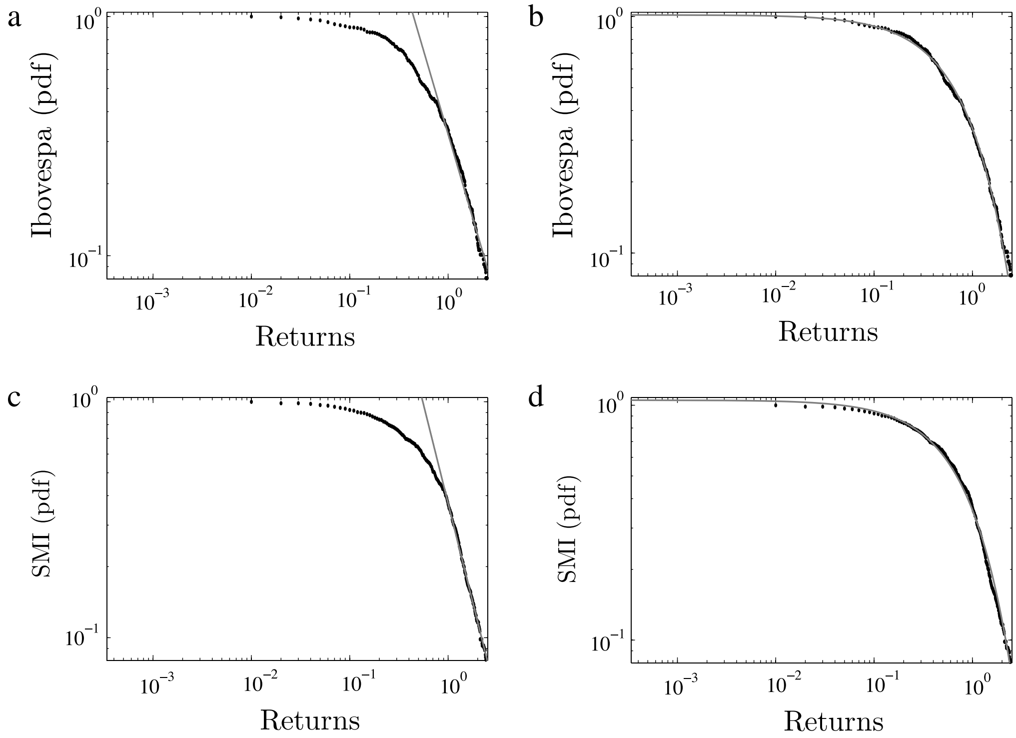

Fig. 1. Probability density function in log–log scale. (a)–(b) Fitting to Ibovespa index with power laws and exponential function, respectively. (c)–(d) Fitting to SMI index with power laws and exponential function, respectively. The black line is the probability density function (pdf) of indices and the gray line is the fitting line.

Fund (IMF), World Bank and Human Development Index (HDI). The market indices are classified in developing and developed markets. A clear contradiction can be observed: the KWSE (Kuwait) index is classified as developed by the WB and as developing by the UN and IMF. In Table 1, all countries classified as developed by the UN and IMF have HDI equal or greater than 0.8.

Each market index is represented by the pdf of its return series. Fig. 1 shows the fitting adjustment of two indices, SMI and Ibovespa, in the log–log scale. The mean squared error (MSE) of the fitting generated by the exponential function and by the power laws [20] can be observed in Table 1. For all the analyzed indices, the exponential function adheres better to the probability density function of the indices than the power laws. This behavior occurs because the power law adheres better in the tail of the pdf [9,21], while the exponential function fits the whole range of the pdf [14]. However, if just the end of the tail is considered (extreme events), the power law fit is very accurate and can describe well this region of the pdf [21].

A correlation between volatility (Eq. (6)) and the B coefficient (Eq. (2)) is observed. Fig. 2 shows that when the volatility decreases, the B coefficient increases. The financial risk of a given market over a specific period of time can be quantified by the volatility and by the B coefficient. In other words, these two features measure the instability/fluctuation of the markets.

As stated in Eq. (4), the expression $B \times \textit{Volatility}$ tends to be a constant. After an experimental study, the mean value of that constant was 1.01 (0.06), as shown in Fig. 3; the standard deviation is within parentheses.

A clustering algorithm can be used to form two distinct groups: one that represents developed markets and another one that represents developing markets. Fig. 4 shows a plot where each one of the 34 indices is represented by a point. The k-means clustering algorithm obtained the following two centroids: $(1.0640, 0.9642)$ for developing markets and $(0.9464, 1.0710)$ for developed markets. Based on these centroids, the whole set of developed markets are correctly classified. However, five developing market indices are incorrectly classified: Argentina (Merval), Mexico (IPC), Morocco (MASI) and Brazil (IBRX 50 and Ibovespa). Thus, the overall hit rate is 85.3%, assuming that the IMF and UN classifications are the ground truth. But, when the clustering is compared with the WB classification, the hit rate decays to 82.3%, because Kuwait is classified as a developed country. When the Kuwait index is removed (it has different classification depending on the international organization), the overall hit rate decreases by approximately 0.5 percentile points with respect to the UN and IMF classifications and increases by 2.5 percentile points with respect to the WB classification.

5. Conclusions

Traditional organizations, such as the United Nations (UN), International Monetary Fund (IMF) and World Bank (WB), use multiple indicators to group countries. This paper proposed a method to cluster countries that uses only two attributes

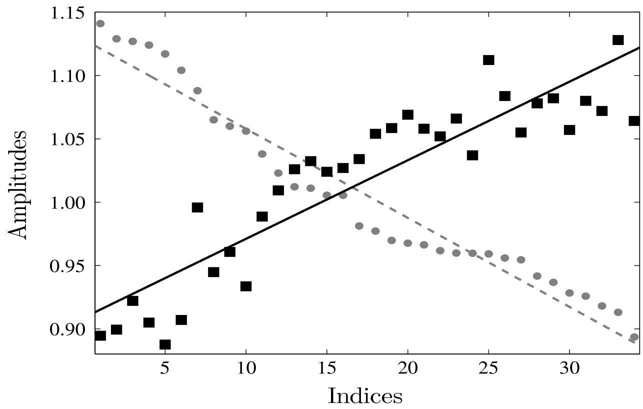

Fig. 2. Amplitude of the volatility and the B coefficient for each one of the 34 considered market indices. The gray dots are the volatilities of the indices and the black squares are the B coefficients. The dashed line and the solid line are the linear fitting of the volatilities and B coefficients, respectively.

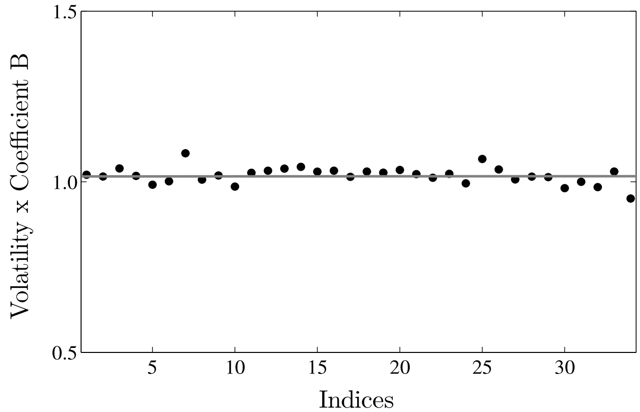

Fig. 3. The dots are the product $B \times Volatility$ for each market and the solid line represents the mean of these values.

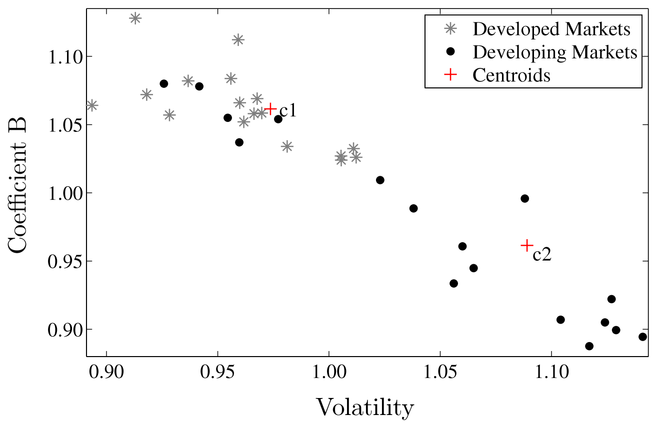

Fig. 4. Each market is represented by a pair (Volatility, B coefficient). The stars represent the developed markets, the dots represent the developing markets and the crosses are the centroids.

(B coefficient and volatility). The method is based on ideal gas theory and treats the stock market as a gas composed of particles (stocks or agents) that have a random movement. These attributes are extracted from the probability density function of the return series of world indices. This pdf describes the dynamics of the fluctuations of the financial markets and it is well adjusted by exponential functions. Using the Maxwell–Boltzmann distribution, a correlation between the dynamics of the stock market and the behavior of an ideal gas was established. Thus, the volatility of a financial system or stock market corresponds to the temperature of an ideal gas — the larger the market fluctuation, the higher its temperature and the higher the volatility of the return series.

The volatility and the B coefficient can be used to characterize the movement of the markets. These two features are used to classify the countries in developed and in developing economies. First, two groups are formed based on the k-means clustering algorithm. Next, the query countries are attributed to the nearest cluster. The results obtained by the proposed method were compared with United Nations (UN), International Monetary Fund (IMF) and World Bank (WB) classifications. The countries classified as developed by the proposed method have the same classification in the UN, IMF and WB. The proposed method found the same class defined by the UN and IMF in 29 out of 34 indices. When compared with WB classifications, this number decreases to 28 indices. However, the acquisition procedure of the features used by the UN, IMF and WB is a difficult, expensive and often inaccurate task. On the other hand, the features proposed in these work are easy to calculate and presented promising results.

The B coefficient of the exponential function reflects the mean behavior of the pdf and the volatility is proportional to the mean dispersion of the data in the time window observed. Thus, depending on the dynamics of the country, if the size of the time windows is very large, the B coefficient and the volatility can not capture the present situation of the country.

The correlation between volatility and B coefficient was found only in daily records of a short time window (January 2008–January 2010) [14]. As future works, other window sizes should be investigated to test if the behavior is the same.

References

[1] M. Gligor, M. Ausloos, Convergence and cluster structures in EU area according to fluctuations in macroeconomic indices, Journal of Economic Integration 23 (2008) 297–330.
[2] W. Bank, How we classify countries, Tech. rep., World Bank, 2010.
[3] I.M. Funds, Country composition of world economic outlook groups, Tech. rep., International Monetary Funds, 2010.
[4] U. Nations, Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings, Tech. rep., United Nations, 2011.
[5] U. Nations, International human development indicators, Tech. rep., United Nations, 2010.
[6] B. Mandelbrot, The variation of certain speculative prices, The Journal of Business 36 (4) (1963) 394–419.
[7] B. Mandelbrot, New methods in statistical economics, Journal of Political Economy 61 (1963) 421–440.
[8] E.F. Fama, Problems with fitting to the power-law distribution, Journal of Business 38 (1965) 34–105.
[9] R.N. Mantegna, E.H. Stanley, Scaling behaviour in the dynamics of an economic index, Nature 376 (6535) (1995) 46–49.
[10] P. Gopikrishnan, V. Plerou, L.A. Nunes Amaral, M. Meyer, H.E. Stanley, Scaling of the distribution of fluctuations of financial market indices, Physical Review E 60 (5) (1999) 5305–5316.
[11] R.N. Mantegna, Levy walks and enhanced diffusion in Milan stock exchange, Physica A: Statistical Mechanics and its Applications 179 (2) (1991) 232–242.
[12] H.E. Stanley, V. Plerou, X. Gabaix, A statistical physics view of financial fluctuations: evidence for scaling and universality, Physica A 387 (2008) 3967–3981.
[13] H.E. Stanley, L. Amaral, P. Gopikrishnan, V. Plerou, M.A. Salinger, Application of computational statistical physics to scale invariance and universality in economic phenomena, Computer Physics Communications 146 (1) (2002) 84–92.
[14] P.S.G. de Mattos Neto, D. Silva, T. Ferreira, G.D.C. Cavalcanti, Market volatility modelling for short time window, Physica A: Statistical Mechanics and its Applications 390 (20) (2011) 3444–3453.
[15] R. Duda, P. Hart, D. Stork, Pattern Classification, Wiley, 2001.
[16] F. Reif, Fundamentals of Statistical and Thermal Physics, in: Fundamentals of Physics, Waveland Press, 2008.
[17] A.C. Silva, R.E. Prange, V.M. Yakovenko, Exponential distribution of financial returns at mesoscopic time lags: a new stylized fact, Physica A: Statistical Mechanics and its Applications 344 (2004) 227–235.
[18] G. Casella, R. Berger, Statistical Inference, in: Duxbury Advanced Series in Statistics and Decision Sciences, Thomson Learning, 2002.
[19] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, 2000.
[20] D.H. Kim, M.Y. Cha, J.W. Lee, The persistence probability and the price–price correlation functions in the Korean stock market, Computer Physics Communications (2011) 243–244.
[21] B. Podobnik, D. Horvatic, A.M. Petersen, H.E. Stanley, Cross-correlations between volume change and price change, Proceedings of the National Academy of Sciences of the United States of America 106 (52) (2009) 22079–22084.