Correlation of worldwide markets' entropies

José A.O. Matos[1,2,3], Sílvio M.A. Gama[1,3], Heather J. Ruskin[4], Adel Sharkasi[4], Martin Crane[4]

[1] Centro de Matemática da Universidade do Porto, Edifício dos Departamentos de Matemática da FCUP, Rua do Campo Alegre 687, 4169-007 Porto, Portugal

[2] Grupo de Matemática e Informática, Faculdade de Economia da Universidade do Porto, Rua Roberto Frias, 4200-464 Porto, Portugal

[3] Departamento de Matemática Aplicada, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal

[4] School of Computing, Dublin City University , Dublin 9, Ireland

Abstract

The goal of this study is the analysis of financial data series from worldwide stock market indices. We have expanded the scope of previous work on the PSI-20 index, since results seemed to provide a basis for a wider ranging study of coherence and entropy. This broader study, both in a geographical and technical sense, explores several measures of disorder, using methods from high frequency analyses such as Detrended Fluctuation Analysis and Hurst exponents, with eigenvalues of the covariance matrix used as a measure of synchronization.

The time-scale dependence of the referred measures demonstrates the relevance of entropy measures in distinguishing the several characteristics of market indices: "effects" include early awareness, patterns of evolution as well as comparative behaviour distinctions in emergent/established markets.

One of the interesting outcomes is that, in spite of the results showing differences between known emergent markets and established ones, we found convergence of entropy behaviour in recent years among the worldwide markets studied. A plausible explanation for this phenomenon is the progressive globalization of financial markets.

1 Introduction

1.1 Goals

The goal of this study is the analysis of stock exchange world indices searching for signs of coherence and/or synchronization across the set of studied markets.

We have expanded the scope of previous work on the PSI-20 (Portuguese Standard Index), since results there [9] seemed to provide a basis for a wider ranging study of coherence and entropy.

With that purpose we applied econophysics techniques related to measures of "disorder"/complexity (entropy) and a newly proposed [12] generalization of Detrended Fluctuation Analysis. As a measure of coherence among a selected set of markets we have studied the eigenvalues of the correlation matrices for two different set of markets, exploring the dichotomy represented by emerging and mature markets.

1.2 Historical motivation

Entropy

The early notion of entropy as a measure of disorder comes from the work of Claussius, in which entropy provides a way to state the second law of Thermodynamics (as well as a definition of temperature). Boltzman extended the idea further giving it a central role in statistical physics. Here entropy is a measure of system multiplicity and can be visualized in terms of disorder.

In 1948 Shannon[7] gave a new meaning to entropy in the context of Information Theory, relating entropy with the absence/presence of information in a given message.

Entropy, one of the early ideas behind thermodynamics that later led the way to the emergence of statistical physics has been shown to be pervasive and perhaps surprisingly crossed disciplinary boundaries (to pure mathematics) giving an easier interpretation to the previously defined concept of topological entropy. The influence of thermodynamics was such that it lent its name to the thermodynamical formalism by the hands of Rufus Bowen and David Ruelle.

The theoretical concept proves to be rich and active as demonstrated in the late 80's when Constantino Tsallis [8] introduced the concept of non-extensive entropy, generalizing further the "traditional" concept of entropy.

Fractional Brownian motion/ Econophysics

A first theory of stock-market fluctuations was proposed by Bachelier [1], five years before Einstein's famous paper on Brownian Motion [2], in which Einstein derived the partial differential heat/diffusion equation governing Brownian motion and made an estimate for the size of molecules. In 1900 Bachelier studied the Paris Stock Exchange in his PhD thesis introducing Brownian motion to describe the evolution of the financial assets.

Bachelier gave the distribution function for what is now known as the Wiener stochastic process – the stochastic process that underlies Brownian Motion – linking it mathematically with the diffusion equation. The probabilist Feller originally referred to this as the Bachelier-Wiener process.

In the 50's Hurst while analyzing hydrological flows, proposed a single exponent to characterize time variation. As can be seen in Section 3.2 this approach is a generalization of Brownian motion later called fractional Brownian motion.

In the 60's Mandelbrot [4] pointed out that the distributions of price differences are not Gaussian due to the so called *fat-tails*. In contrast, these distributions display power-law decay in the tails and this is related to the fractal nature of financial data.

In problems with similar data characteristics [6], for example, while studying the DNA patterns and characterizing them, presents Detrended Fluctuation Analysis (DFA) as a method to estimate the Hurst exponent.

2 Data

2.1 Features

The data used in this study was taken daily for a set of worldwide market indices, enumerated in Section 2.2. As it is usual in this kind of analysis we base our results on the study of returns

$$\eta_i = \log \frac{x_i}{x_{i-1}},$$

where η_i is the return at time step i.

2.2 List of market indexes

Abbrev.	Index Name	Country
^AEX	AEX General	Netherlands
^AORD	All Ordinaries	Australia
^ATX	ATX	Austria
^BFX	BEL-20	Belgium
^BSESN	BSE 30	India
^BVSP	Bovespa	Brazil
^CCSI	CMA	Egypt
^CSE	All Share	Sri Lanka
^FCHI	CAC 40	France
^FTSE	FTSE 100	United Kingdom
^GDAXI	DAX	Germany
^GSPC	500 Index	United States
^GSPTSE	S&P TSX Composite	Canada
^HSI	Hang Seng	Hong Kong
^IPSA	IPSA	Chile
^ISCI	ISEC Small Cap	Ireland
^ISCT	ISEC Small Cap Techno	Ireland
^ISEQ	Irish SE Index	Ireland
^IXIC	Nas/NMS Composite (Nasdaq)	United States
^JKSE	Jakarta Composite	Indonesia
^KFX	KFX	Denmark
^KLSE	KLSE Composite	Malaysia
^KS11	Seoul Composite	South Korea
^KSE	Karachi 100	Pakistan
^MERV	MerVal	Argentina
^MIBTEL	MIBTel	Italy
^MTMS	Moscow Times	Russia
^MXX	IPC	Mexico
^N225	Nikkei 225	Japan
^NZ50	NZSE 50	New Zealand
^OSEAX	OSE All Share	Norway
^PSI20	PSI 20	Portugal
^PSI	PSE Composite	Philippines
^PX50	PX50	Czech Republic
^SETI	SET	Thailand
^SMSI	Madrid General	Spain
^SSEC	Shanghai Composite	China
^SSMI	Swiss Market	Switzerland
^STI	Straits Times	Singapore
^SXAXPI	Stockholm General	Sweden
^TA100	TA-100	Israel
^TWII	Taiwan Weighted	Taiwan
^XU100	ISE National-100	Turkey

3 Techniques

The econophysics techniques applied in this work are twofold: measures of "disorder"/complexity and measures of coherence.

The measures of "disorder" and complexity are the entropy (as defined by Shannon information theory) and the fractional Brownian motion. We study the temporal dependency of entropy for a representative set of markets and the time-scale dependency of the Hurst exponent obtained through DFA.

We use the covariance matrix as a measure of coherence among a close set of related markets. In this work we distinguish between mature and emerging markets, and look to the time dependency of the ratio between several of the most significant eigenvalues of the covariance matrix.

3.1 Entropy

The definition of entropy is the following:

Definition Let X be a discrete random variable on a finite set $\mathcal{X} = \{x_1, ..., x_n\}$, with a probability distribution function $p(x) = \mathrm{P}(X = x)$. The entropy $H(X)$ of X is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

Since the original data is best described by a continuous random variable instead of a discrete random variable as the previous formula requires, we had to discretize the data. We split the distance between the maximum and minimum for all studied markets into 50 bins with the same size. These bins become the partition used in previous definition.

This partition allowed us to assign a “letter” to each day, from this we have evaluated 5-letter sequences. This corresponds to a week in trading time.

It should be noted that results are robust to the choice of the total number of bins (the size of our alphabet). That is, a different choice of the number of partitions yields similar results.

In order to enhance the time dependence of results we have evaluated the entropy of the set for for periods of 100 trading days (roughly corresponding to half a year). This method allowed us to describe the time dependence of the results.

Results

The results displayed in Figure 1 show interesting coherence (produced entropy) after 1997 as compared with previous periods. Higher entropy implies less predictability, which seems to be the case for all markets after 1997.

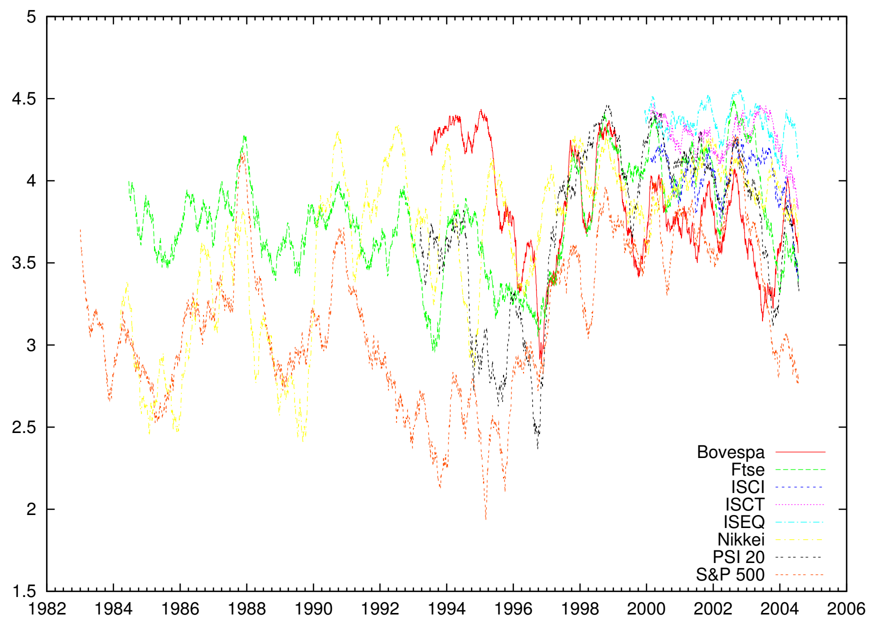


Figure 1: Weekly entropy for various market indexes

The remarking feature of this graphic is that it affects mature as well as developing markets alike.

3.2 Detrended Fluctuation Analysis (DFA)

Fractional Brownian motion

Fractional Brownian motion (fBm) is a well-known stochastic process where the second order moments of the increments scale as

$$E\{(X(t_2) - X(t_1))^2\} \propto |t_2 - t_1|^{2H} \tag{1}$$

with $H \in [0, 1]$. The Brownian motion is then the particular case where $H = 1/2$.

The exponent H is called the Hurst exponent. If $H < 1/2$, then the behaviour is anti-persistent, that is, deviations of one sign are generally followed by deviations with the opposite sign. The limiting case $H = 0$, corresponds to white noise, where fluctuations at all frequencies are equally present.

If $H > 1/2$, then the behaviour is persistent (smooth), i.e., deviations tend to keep the same sign. The limiting case $H = 1$, reflects $X(t) \propto t$, a smooth signal.

While motivation for fBm was the fat-tail characteristic of real price distributions, this $H-$threshold for persistent/anti-persistent behaviour is useful in terms of determining when trends break down.

Method description

The DFA technique consist in dividing a random variable sequence $X(n)$, of length N, into N/t non-overlapping boxes, each containing t points [6]. Then, the linear local trend $z(n) = a\,n + b$ in each box is defined to be the standard linear least-square fit of the data points in that box. The detrended fluctuation function F is then defined by:

$$F^2(t) = \frac{1}{t} \sum_{n=kt+1}^{(k+1)t} |X(n) - z(n)|^2 , \qquad k = 0, \cdots , \frac{N}{t} - 1.$$

Averaging $F(t)$ over the N/t intervals gives the fluctuation $< F(t) >$ as a function of t. If the observable $X(n)$ are random uncorrelated variables or short-range correlated variables, the behaviour is expected to be a power law

$$\langle F(t) \rangle \sim t^H.$$

DFA has the advantage over standard variance analysis of being able to detect long-term dependence in non-stationary time series. Additionally, the advantages of DFA in computation of H over other techniques (for instance, the Fourier transform) are:

1. inherent trends are avoided at all t scales;
2. local correlations can be easily probed.

It should be noted that DFA is a crude measure, especially as it is sensitive to non-normality of data.

On a practical note it should be said that it is not enough to obtain an exponent from the detrended function, but is necessary also to check the quality of the fit. We have used the correlation coefficient (r) as a crude measure of the goodness of the fit. All results showed a high value of r giving us some confidence in the results obtained through DFA.

Time and Scale DFA (TSDFA)

One possible generalization of DFA involves the same scheme used in the previous study of entropy. We fix the time window to be a given size for which we evaluate the Hurst exponent obtaining thus a function of time $H(t)$. For PSI-20 [9] results are presented for different sizes of the enclosing window (scale), corresponding to 100, 200 and 400 trading days.

The next obvious generalization after this step is described and fully explored in [12], where we propose a new method called Time and Scale DFA (TSDFA) in which the Hurst exponent is dependent on both time and scale, $H(s, t)$. $H(s, t)$ is evaluated taking the sub-series $[t - s/2; t + s/2]$ and using DFA.

It is interesting to note that this method is akin to Continuous Wavelet Transform [10] in its dependence on time and scale (both tools provide redundant information to some extent).

Results

We have chosen different markets, in different stages of maturation, to illustrate the results from applying this method, but have use the same time interval to facilitate the comparison of markets. (See Figures 2-7.)

An overview of the graphics allows us to distinguish immediately mature and developing markets, and in some cases we can even identify some sort of "phase transition" the most striking case being the Canadian market (Figure 7).

Interestingly these graphics show stripes that appear in all scales, and they point events, for instance that corresponding to the terrorist attack of September 11th shows in all graphics and across all scales.

Another tendency illustrated here is the maturation of markets with time. We see values of values of $H(s, t)$ closer to 0.5 for more recent values of t.

3.3 Covariance matrices

The work explored here was developed in [11]. We use the covariance matrix to study the coherence of various types of assets. For this study we have considered the traditional distinction between mature and emerging markets.

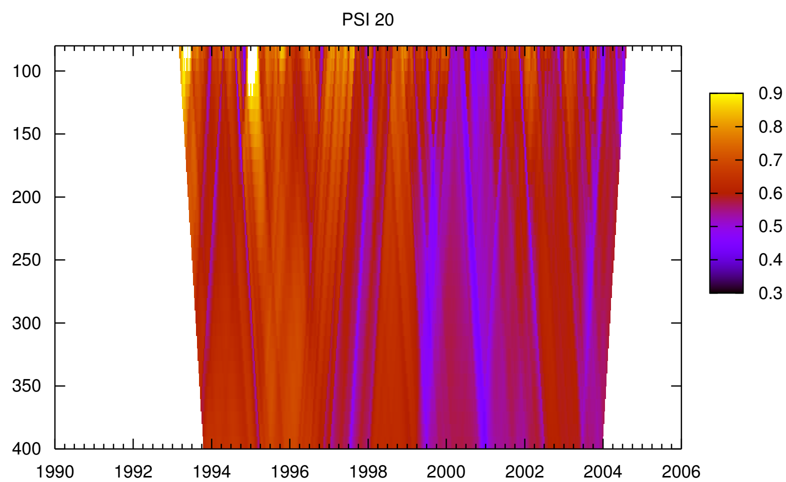


Figure 2: PSI-20, Portugal. The y's axis represents the scale and the x's axis represents the time. This contour plot shows the Hurst exponent obtained applying the TSDFA.

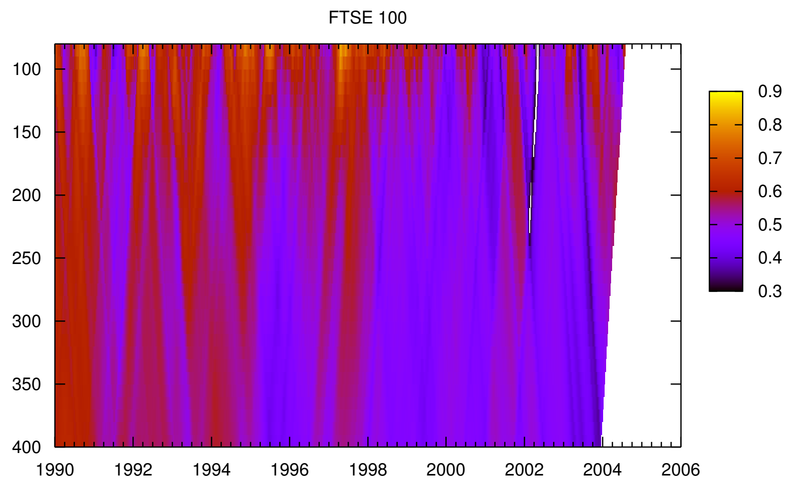


Figure 3: FTSE, UK

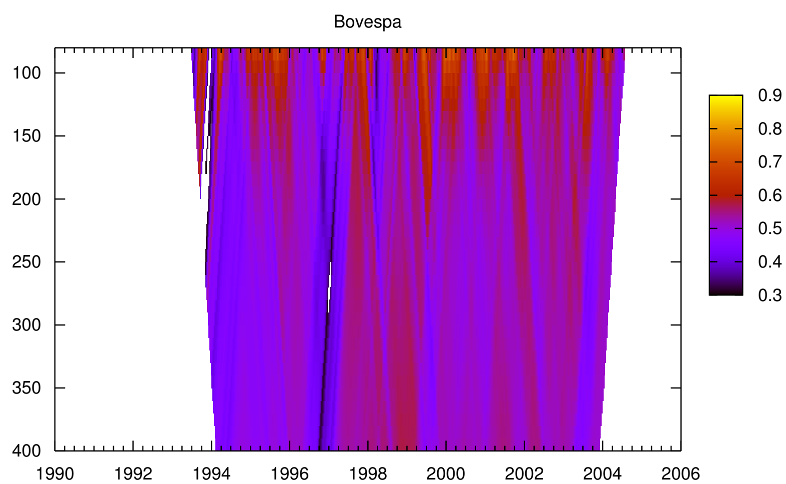


Figure 4: Bovespa, Brazil

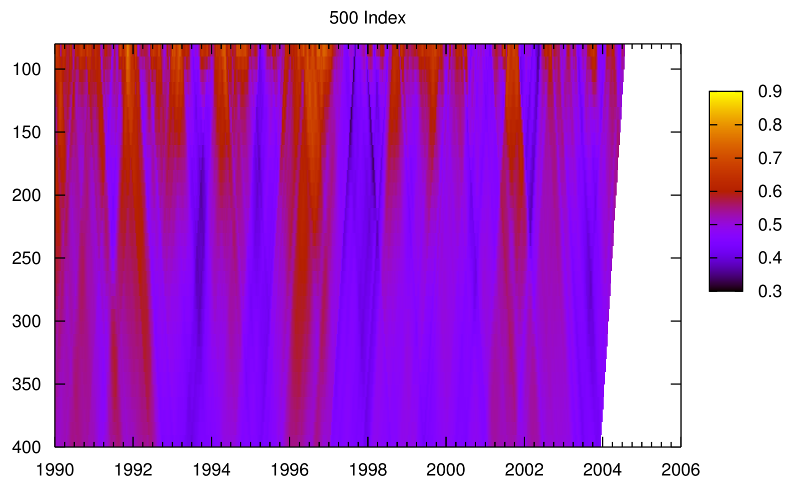


Figure 5: S&P 500, USA

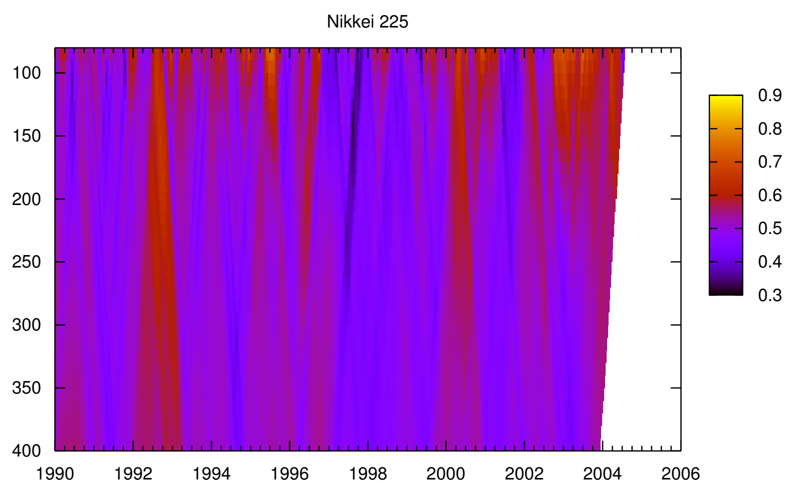


Figure 6: Nikkei, Japan

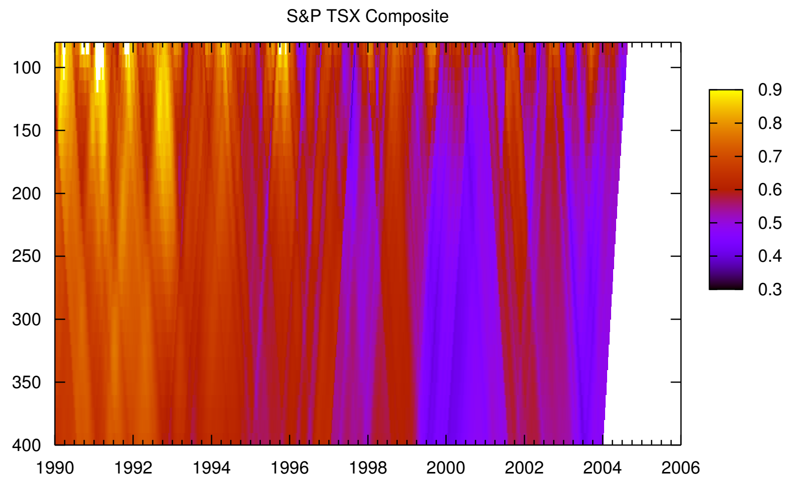


Figure 7: S&P TSX Composite, Canada

The covariance matrix with variable weights is given by:

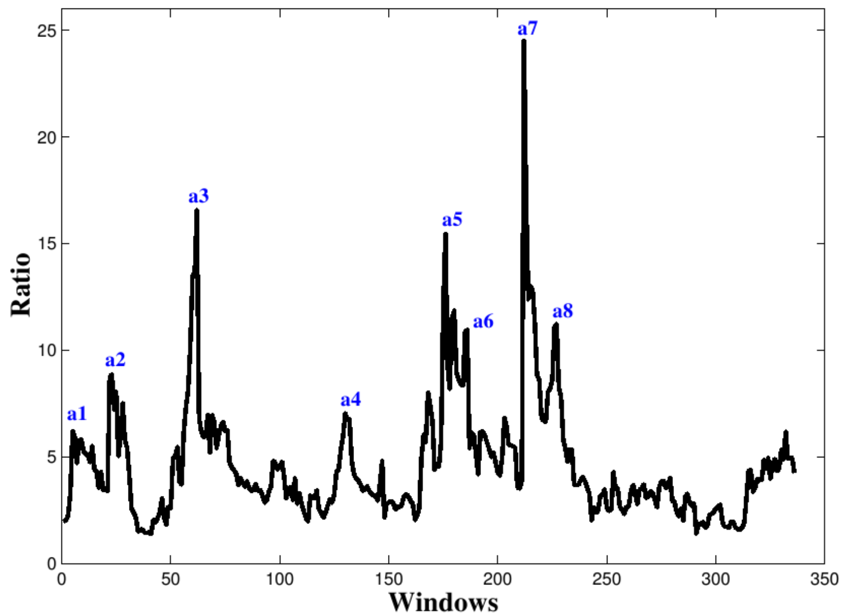
$$\sigma_{ij}^{T}(t) = \frac{\sum_{s=0}^{T} W_{T-s} r_{i,t+T-s} r_{j,t+T-s}}{\sum_{s=0}^{T} W_{T-s}}.$$

where:

- $r_{i,t}$ return of asset i at time t;
- W_s weight for delay s;
- T length of interval.

Tradicional usage considers $W_i = R^i$, with $0 < R < 1$, resulting $\sum_{s=0}^{T} W_{T-s} = \frac{R^T}{1-R^T}$. This case corresponds to a geometric mean. The cut-off period T corresponds to a period where R^T becomes negligible so we can truncate the series at that point.

Typical values for those type of data are $R = 0.9$ and $T = 20$ and used. Weekly periods have been used to estimate the returns.



(a) **Original Return series for emerging markets.**

Figure 8: Evolution of for $\frac{\lambda_1}{\lambda_3}$ emerging markets

Results

The results presented in this paper are illustrative of those discussed in the original paper [11]. As an example, in Figure 8, we represent the ratio between the first and the third most important eigenvalues $(\frac{\lambda_1}{\lambda_3})$ for a given set of emerging markets. The same analysis applies for mature markets, see Figure 9.

Again interest lies in the fact that spikes in Figures 8 and 9 correlate with real events as displayed in Tables 1 and 2 respectively.

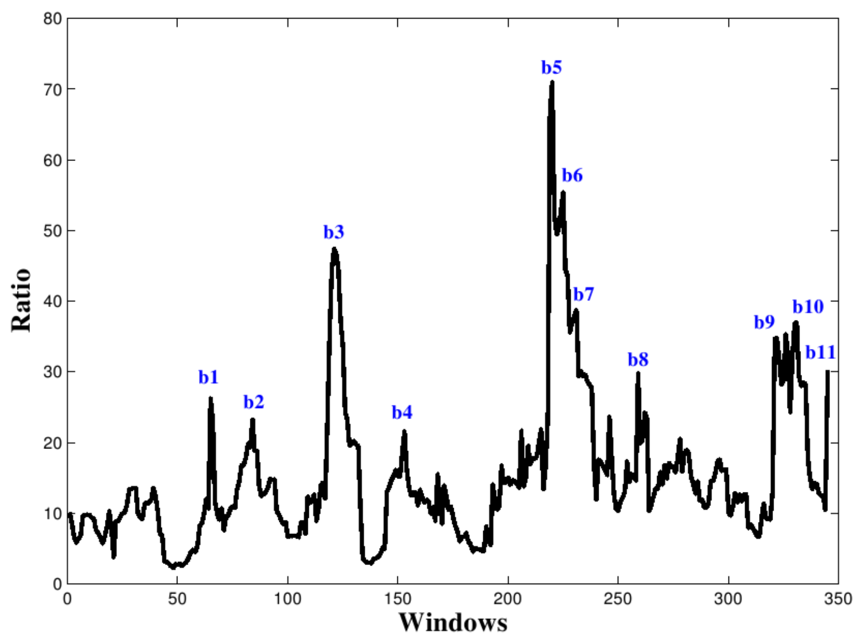
4 Conclusions

We have focused on the aspect of time dependence for several econophysics techniques, applied to markets, categorized as emerging or mature and subject to diverse levels of disorder or volatility in their financial series. The out-

Mark	Window No.	Last week included	Events
a1	5	first week of 7/1997	Asian Crash
a2	23	second week of 11/1997	Asian Crash
a3	62	fourth week of 8/1998	Global Crash
a4	130	second week of 1/2000	
a5	176	second week of 12/2000	Effects of DotCom Crash
a6	186	second week of 3/2001	
a7	212	second week of 9/2001	September the 11^{th} Crash
a8	227	fourth week of 1/2002	

a: Emerging Markets.

Table 1: Table of events (emerging)



(b) **Original Return series for mature markets.**

Figure 9: Evolution of $\frac{\lambda_1}{\lambda_3}$ for mature markets

Mark	Window No.	Last week included	Events
b1	65	first week of 9/1998	Global Crash
b2	84	fourth week of 12/1998	Global Crash
b3	121	third week of 10/1999	Last October in 20^{th} Century
b4	153	second week of 6/2000	DotCom Crash
b5	220	second week of 9/2001	September the 11^{th} Crash
b6	225	first week of 11/2001	Effects of 9/11 Crash
b7	231	second week of 12/2001	Effects of 9/11 Crash
b8	259	first week of 5/2002	The Stock Market Downturn
b9	322	first week of 10/2003	
b10	331	first week of 12/2003	General Threat Level Raised
b11	345	third week of 3/2004	Madrid Bomb

b: Mature Markets.

Table 2: Table of events (mature)

come shows clear synchronization of world markets, observed in the weekly entropy.

Explanation of the dynamical nature of markets is enhanced through time-scaled detrended fluctuation analysis (TSDFA) where we investigate both time and scale dependence of market behaviour. As discussed in a companion paper [12] this technique calls for a better classification for markets, where the simple dichotomy between mature and emerging markets supply a "first order" approach to classification.

Most markets studied show that maturity is developing as evidenced by the Hurst exponent values decrease over time, getting close to 0.5 the random walk limit.

Acknowledgments

One of authors (JAOM) would like to thank ESF (European Science Foundation) for COST action P10 - STSM 00421, that made possible a visit to Dublin City University where part of this work was started.

References

[1] L. Bachelier, Ann. Sci. Ecole Norm. S., III : 17 (1900) 21.

[2] A. Einstein, Ann. Phys-Berlin, 17 (1905) 549.

[3] H. Hurst, Trans. Am. Soc. Civ. Eng., 116 (1951) 770.

[4] B. Mandelbrot, J. Bus., XXXVI (1963) 392.

[5] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics, Cambridge University Press, Cambridge, 2000.

[6] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Phys. Rev. E, 49 (1994) 1685.

[7] C. Shannon, Bell Syst. Tech. J., 27 (1948) 379.

[8] C. Tsallis, J. Stat. Phys., 52, (1988) 479.

[9] J.A.O. Matos, S.M.A. Gama, H. J. Ruskin, J.A.M.S Duarte, Phys. A, 342 (2004) 665.

[10] D. Percival and A. Walden, Wavelet methods for time series analysis, Cambridge University Press (2000).

[11] A. Sharkasi, M. Crane, H. J. Ruskin, J. A.O. Matos, Phys. A, to appear (2006).

[12] J. Matos, S. Gama, A. Sharkasi, H. Ruskin, M. Crane (in preparation) (2006)