JOURNAL
OF
APPLIED
QUANTITATIVE
METHODS



# INFORMATION ENTROPY AND OCCURRENCE OF EXTREME NEGATIVE RETURNS[1]

**Daniel Traian PELE**

PhD, Assistant Professor, Department of Statistics and Econometrics,
University of Economic Studies, Bucharest, Romania

**E-mail:** danpele@ase.ro

**Abstract:**

*This study is intended to investigate the connection between the complexity of a capital market and the occurrence of dramatic decreases in transaction prices. The work hypothesis is that such episodes, characterized by sudden and dramatic decreases in transaction prices mostly occur in period of market inefficiency, when the level of complexity reaches a local minimum. In this regard, we introduce a complexity estimator, through differential entropy. The connection between the market complexity level and the appearance of extreme returns is illustrated in a logistic regression model.*

**Key words:** *differential entropy; stock market crash; logistic regression*

## 1. Introduction

The work hypothesis tested in this paper consists in the fact that the phenomenon of capital market crashes is mostly manifested when the market registers a significant decrease in the complexity level as against to the efficient market hypothesis.

The connection between a certain extent of complexity and the efficient market hypothesis is quite obvious: if we assume an efficient market (weak form) then the price of an asset which follows a random walk model, i.e. the series of returns is a white noise process. From the quantitative measurements point of view, a white noise process is the most complex process possible; on the contrary, if the efficient market hypothesis is not respected, than the price is no longer a random walk process and consequently, the market complexity level is lower.

For instance, if the price is purely deterministic process, completely predictable, then we can speak of reaching a minimum complexity level; if however, the price is purely random, completely unpredictable we can speak of a maximum complexity level. Our study is intended to prove that episode such as stock market crashes, characterized by sudden and dramatic decreases in transaction prices mostly occur in periods of market inefficiency, when their complexity level reaches a local minimum level.

Risso (2008) uses entropy as a complexity measurement in order to investigate the hypothesis according to which stock market crashes are associated to low entropy periods. The connection between information efficiency and stock market crashes is the following: if the market is inefficient, thus the information not being reflected instantaneously in prices there can be created trends in price evolution. However, when information reaches investors, the price can resettle, leading to significant crashes.

## 2. Information entropy as a measure of complexity

Entropy is measure of complexity, having numerous applications in physics, information theory, biology, medicine, and economics.

In its classical formulation, entropy can be defined in a discrete space.

If we have a discrete random variable X, with the distribution $X:\begin{pmatrix} x_1......x_n \\ p_1......p_n \end{pmatrix}$,

where $p_i = P(X = x_i)$, $0 \le p_i \le 1$ and $\sum_i p_i = 1$, then the Shannon information entropy is defines as:

$$H(X) = -\sum_i p_i \log_2 p_i \,.$$

We notice that for a uniform distribution, we obtain the maximum value for entropy: $H(X) = -\sum_i (1/n)\log_2(1/n) = \log_2 n$; the minimum value is obtained for a distribution like $X:\begin{pmatrix} x_1......x_n \\ 1.........0 \end{pmatrix}$, $H(X) = 0$.

In other words, larger values of the entropy are obtained for situations with high certainty, while smaller values are associated with low certainty situations.

The application in case of stock markets is immediate.

Let $r_t = p_t - p_{t-1} = \log P_t - \log P_{t-1}$ be the return associated to an asset and $s_t = \begin{cases} 1, r_t > 0 \\ 0, r_t < 0 \end{cases}$ a variable which defines the „bull-bear" statuses.

Then, for a certain period of time, we can define information entropy of a sequence of 0 and 1: $H = -p\log_2 p - (1-p)\log_2(1-p)$, where $p = P(s_t = 1) = 1 - P(s_t = 0)$.

## 3. The theoretical model of the connection between complexity and stock exchange market crashes

Risso (2008) uses this measure of complexity to investigate the hypothesis according to which stock exchange market crashes are associated especially to low entropy periods.

In fact, entropy can be looked upon as a measure of information efficiency in capital markets: if the market is efficient in its weak form, then the price is a random walk process, and bull and bear market situations are consequently equally probable. In terms of information entropy, market efficiency is equivalent to the situation of maximum entropy, of maximum complexity. On the contrary, situations characterized by ascendant or descendant trends, situations when price predictability is hinted, are characterized by a lower complexity level, and thus, a smaller entropy value.

In order to verify this hypothesis, we can estimate a logistic model:

$\log\frac{P(y^*=1)}{1-P(y^*=1)}=\beta_0+\beta_1 H$ or equivalently, $P(y_t^*=1)=\frac{\exp(\beta_0+\beta_1 H_t)}{1+\exp(\beta_0+\beta_1 H_t)}$.

In the equation above, we have:

- $y_t^*=1$ for the smallest 1% of the returns ( $y_t^*=\{1\,|\,r_t<r_t^*, P(r_t<r_t^*)=0.01\}$ )
- $H_t$ Is market information entropy at moment t.

The issue that is imposed is how to measure information entropy. The methodology used by Risso (2008) is the following:

- For a certain time interval *T*, we define the values $s_t=\begin{cases}1, r_t>0\\ 0, r_t<0\end{cases}$, obtaining a succession of zeros and ones.
- We define a rolling window $v<<T$ for which we compute the probability $p_v=P(s_t=1)$
- The entropy for the entire time interval will then be $H=-p_v\log_2 p_v-(1-p_v)\log_2(1-p_v)$. Also, the normalized Shannon entropy can be computed as $NH=\frac{H}{\log_2 n}$.

Risso applies this methodology on various time intervals *T* and various windows $v$, obtaining statistically significant results for stock market indices in Russia, Japan, Mexico, Malaysia and the USA.

From our point of view, the methodology should be extended, considering the fact that in the present stage we only take into account information contained in 0 and 1 sequences defined above, without return values for the analyzed period.

In other words, information entropy is computed on the discrete case, ignoring the continuous nature of the distribution of returns.

In the followings, we propose a methodology which takes into account this aspect.

## 4. Differential entropy

Unlike the case of a discrete random variable, the entropy of a continuous random variable is much harder to quantify.

For instance, if *X* is a random variable with a density function of $f(x)$ probability, then we can define, by analogy with Shannon information entropy, the differential entropy:

$H(f) = -\int_A f(x)\log_2 f(x)dx$, where A is the support set for X.

Unfortunately, the differential entropy doesn't have all the properties of Shannon entropy: it can take negative values and moreover, is not invariant to linear transformations on variables.

Moreover, there are difficulties when it comes to estimating differential entropy in a sample.

We can define a naïve estimator of differential entropy in the following manner (Lorentz, 2009):

- Let *h>0*. For any *i*, there is $x_i \in [(i-1)h, ih)$ so that $f(x_i)h = \int_{(i-1)h}^{ih} f(x)dx$. This process transforms a continuous variable X into a discrete variable $X_h$, with $P(X_h = x_i) = f(x_i)h$.
- In the conditions above, we can define the entropy for $X_h$ as: $H(X_h) = -\sum_i f(x_i)h\log_2(f(x_i)h) = -\sum_i f(x_i)h\log_2 f(x_i) - \log_2 h$.
- In the conditions above we can define the entropy for $X_h$: $H(X_h) = -\sum_i f(x_i)h\log_2(f(x_i)h) = -\sum_i f(x_i)h\log_2 f(x_i) - \log_2 h$.
- We have: $\lim_{h\to 0} H(X_h) = \lim_{h\to 0}\left[-\sum_i f(x_i)h\log_2 f(x_i)\right] - \lim_{h\to 0}\log_2 h = H(f) + \infty$.

Consequently, the naïve estimator of differential entropy does not converge, which can raise serious issues when *h* is close enough to 0.

Nonetheless, we can define the entropy of a function which fulfills certain properties through a transformation called quantization (this process will constitute the subject of a future study).

In reality, most of the times we don't know the expression of the probability density function, this being precisely the case of the series of returns on capital market.

On the other hand, we can estimate this density using a non-parametric approach, such as the kernel density estimation (KDE).

In essence, KDE assumes the discretization of the continuous distribution and then the estimation of a continuous density around each point $x_i$ of the discretized distribution.

Thus, the KDE estimator has the shape $f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$.

In the previous expression, *K* is a real function, with the following properties:

i) $K(x) \geq 0, \forall x \in \mathbf{R}$.

ii) $K(x) = K(-x), \forall x \in \mathbf{R}$.

iii) $\int_{\mathbf{R}} K(x)dx = 1$.

iv) $\int_{\mathbf{R}} xK(x)dx = 0.$

Such a function is called *kernel* and is usually chosen among the known probability density functions. For instance, there are frequently used:

- Uniform kernel: $K(x) = \frac{1}{2}\mathbf{1}_{(|x|<1)}$;
- Epanechnikov kernel: $K(x) = \frac{3}{4}(1-x^2)\mathbf{1}_{(|x|<1)}$;
- Gaussian kernel: $K(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$;
- Triangular kernel: $K(x) = (1-|x|)\mathbf{1}_{(|x|<1)}$

In order to estimate the differential entropy, we estimate the probability density function of returns, using Kernel Density Estimation(KDE).

Basically, our methodology assumes following the next steps:

- Let $r_t$ be a return series for a time period *T*.
- We choose $n = 2^k$ and estimate probability density function through the formula above, obtaining the $f(x_i)$ values for $i = 0,..,n-1$.
- We estimate the differential entropy with the formula:
  $H(X_h) = -\sum_i f(x_i)h\log_2(f(x_i)h) = -\sum_i f(x_i)h\log_2 f(x_i) - \log_2 h$,
  where $h = 1/n$.

In the actual estimation of differential entropy, we considered a Gaussian kernel and chose $n = 2^7 = 128$ (such that the smoothing factor *h*, is not very close to zero).

## 5. Results

In order to assess the level of complexity of Bucharest Stock Exchange, we estimated the value of the differential entropy and those of normalized Shannon entropy for various rolling-windows and sub-periods in the case of the daily BET returns in the period 1997-2011 (January 31st 2011 is the last value of the series).

Also, we have estimated the following model of logistic regression:

$\log\frac{P(y_t^* = 1)}{1-P(y_t^* = 1)} = \beta_0 + \beta_1 H_t$ or $P(y_t^* = 1) = \frac{\exp(\beta_0 + \beta_1 H_t)}{1+\exp(\beta_0 + \beta_1 H_t)}$.

In the equation above we have:

- $y_t^* = 1$ for the smallest 1% of returns ( $y_t^* = \{1 \mid r_t < r_t^*, P(r_t < r_t^*) = 0.01\}$ ). In the case of BET index for the analyzed period, $r_t^* = -0.05516$.
- $H_t$ is the information entropy of the market at moment *t,* measured successively through the differential entropy previously defined and through the normalized Shannon entropy.

Since we are in the situation of choosing one of more models, we will have to use a performance indicator for the logistic regression model.

In general, such an indicator is defined by comparing the verosimility function of the model with the verosimility function of the model which excludes the exogenous variable.

Thus we have, *pseudo-* $R^2$, a measurement of model performance (Nagelkerke, 1991): $R^2 = 1 - \exp\{2[\log L(M) - \log L(0)]/n\}$, where $L(M)$ and $L(0)$ are the verosimility functions of the model with and without the exogenous variable.

This indicator cannot be interpreted as the weight of variance explained by the model like in the case of classical regression.

If we rewrite the relation above $-\log(1-R^2) = 2[\log L(M) - \log L(0)]/n$ it can be interpreted as the information surplus brought by the exogenous variable.

Unfortunately, $R^2$ in the case of the logistic regression doesn't reach the value 1 in the perfect model either, which is why an adjustment has been proposed (Nagelkerke, 1991):

$$R^2_{adj} = R^2/[1-\exp(2\log L(0)/n)].$$

**Table 1. Adjusted $R^2$ for the two estimators of complexity**

| | Normalized Shannon Entropy | | | | | Differential Entropy | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $v=1$ | $v=2$ | $v=3$ | $v=4$ | $v=5$ | $v=1$ | $v=2$ | $v=3$ | $v=4$ | $v=5$ |
| ***T=60*** | 0.021 | 0.012 | 0.019 | 0.031 | 0.02 | **0.083** | 0.052 | 0.014 | 0.015 | 0.021 |
| ***T=100*** | 0.031 | 0.026 | 0.035 | 0.036 | 0.009 | 0.046 | 0.032 | 0.008 | 0.006 | 0.022 |
| ***T=150*** | 0.017 | 0.01 | 0.015 | 0.032 | 0.009 | 0.028 | 0.015 | 0.002 | 0.002 | 0.012 |
| ***T=200*** | 0.017 | 0.009 | 0.013 | 0.022 | 0.01 | 0.008 | 0.003 | 0.00 | 0.00 | 0.005 |
| ***T=240*** | 0.001 | 0.001 | 0.004 | 0.011 | 0.007 | 0.001 | 0.00 | 0.001 | 0.002 | 0.00 |

As can be seen in the $R^2_{adj}$ values for the estimated logistic regression models, the best performances are offered by the differential dntropy estimator, for $T = 60$ and $v = 1$.

The fact that the best results in the case of the BET index were obtained for $T = 60$, suggests that the Romanian market doesn't present persistent memory in time, the local temporal context being the predominant one.

Also, estimating complexity using the differential entropy of returns offers better results than the classical Shannon entropy.

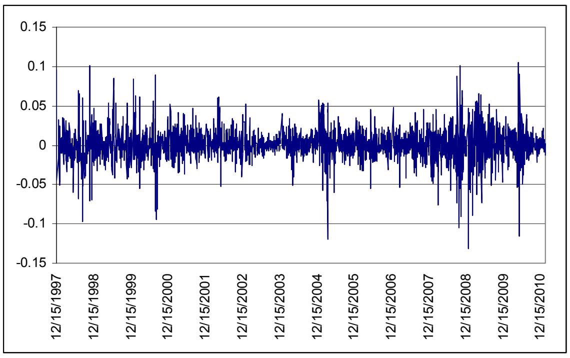


***Fig.1. Daily logreturns of BET Index***

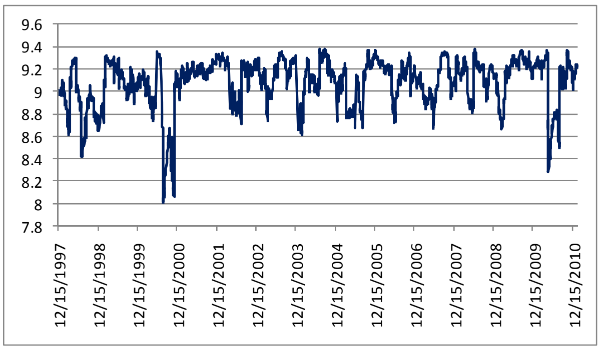


***Fig.2. Differential Entropy(*** $T = 60$, $v = 1$ ***)***

The results of the optimal model estimations, for differential entropy, $T = 60$ and $\nu = 1$ are presented below.

**Table 2. Logistic regression model for differential entropy with $T = 60$ and $\nu = 1$**

| Response Profile | | |
|---|---|---|
| Ordered Value | $y_t$ | Total Frequency |
| 1 | 1 | 32 |
| 2 | 0 | 3234 |

| Model Fit Statistics | | |
|---|---|---|
| Criterion | Intercept Only | Intercept and Covariates |
| AIC | 361.723 | 335.218 |
| SC | 367.814 | 347.400 |
| -2 Log L | 359.723 | 331.218 |

| Testing Global Null Hypothesis: BETA=0 | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 28.5054 | 1 | <.0001 |
| Score | 41.6738 | 1 | <.0001 |
| Wald | 36.0161 | 1 | <.0001 |

| | | | |
|---|---|---|---|
| R-Square | 0.0087 | Max-rescaled R-Square | 0.0833 |

| Analysis of Maximum Likelihood Estimates | | | | | |
|---|---|---|---|---|---|
| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
| Intercept | 1 | 23.5458 | 4.6265 | 25.9016 | <.0001 |
| Differential Entropy | 1 | -3.1368 | 0.5227 | 36.0161 | <.0001 |

| Odds Ratio Estimates | | | |
|---|---|---|---|
| Effect | Point Estimate | 95% Wald Confidence Limits | |
| Differential Entropy | 0.043 | 0.016 | 0.121 |

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 70.8 | Somers' D | 0.500 |
| Percent Discordant | 20.8 | Gamma | 0.547 |
| Percent Tied | 8.4 | Tau-a | 0.010 |
| Pairs | 103488 | c | 0.750 |

| Hosmer and Lemeshow Goodness-of-Fit Test | | |
|---|---|---|
| Chi-Square | DF | Pr > ChiSq |
| 8.6280 | 8 | 0.3746 |

From analyzing the estimation results, we can observe that entropy has a negative influence on the probability of stock market crashes. In fact, an increase of one unit in entropy leads to a decrease of approximately 95% of the chances of crash appearance.

Considering all other indicators which verify the significance of the regression model, we conclude that it is possible to use the differential entropy behavior to anticipate a possible crash.

To illustrate the previous statement, we have studied the behavior of entropy around the main crashes of BET.

**Table 3. The main crashes at Bucharest Stock Exchange**

| Date | Logreturn |
|---|---|
| 1/7/2009 | -0.13117 |
| 3/28/2005 | -0.11902 |
| 5/25/2010 | -0.11612 |
| 10/10/2008 | -0.10454 |

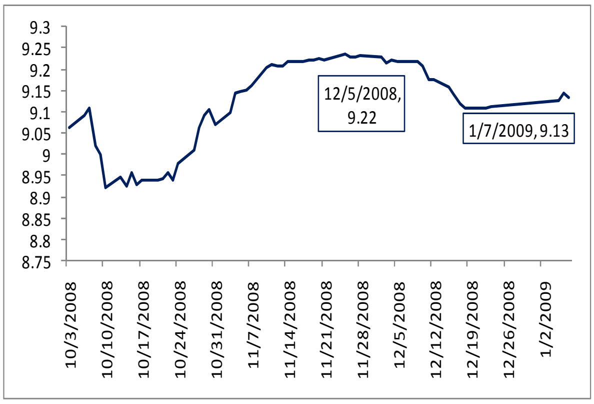


**Fig. 3 The crash from 7 january 2009**

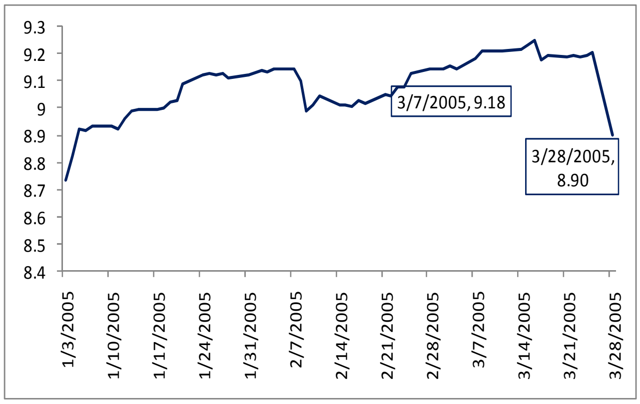


**Fig.4 The crash from 28 march 2005**

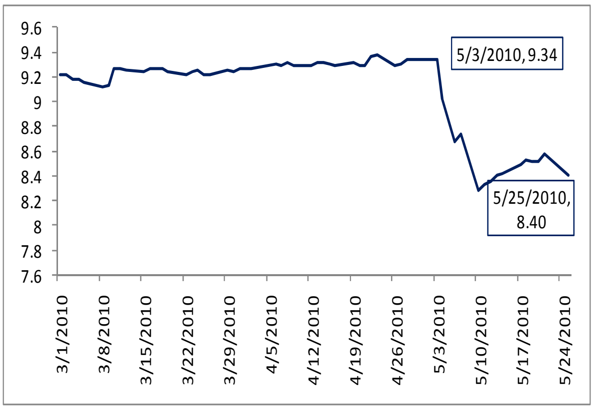


**Fig. 5 The crash from 25 may 2010**

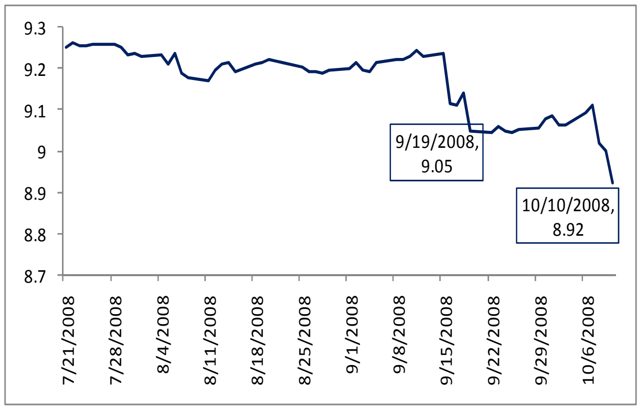


**Fig. 6 The crash from 10 october 2008**

After analyzing the entropy behavior before crash incidence we can observe that such an event is preceded by an entropy local minimum, but the transmission is not instantaneous. This can also be an effect of the Romanian market efficiency deviation from the efficient market hypothesis.

## 6. Conclusions

Capital market information efficiency is a subject intensely debated in the last few years, especially due to the present economic and financial crisis. The connection between capital market efficiency and the predictability of stock market crashes can be illustrated through a complexity measurement, starting from the hypothesis that large decrease episodes in the price of an asset precisely exploit a certain context of efficiency weakening.

In this study, we have analyzed the connection between capital market complexity and stock market crash predictability using the differential entropy of returns as measurement of complexity. The results of the estimated model show that this measurement of complexity generates better results than the classical estimator of Shannon entropy, as it makes it possible to go from a discrete estimator of entropy to a continuous one.

The analysis performed in the case of BET index of Bucharest Stock Exchange has led to the conclusion that the main depreciations of BET have been preceded by episodes of dramatic decreases in the entropy level. In this respect, the entropy behavior can be constituted in a early warning system on the possible negative evolutions of the capital market.

Unfortunately, this method has certain limits, because the entropy differential estimator does not have the required convergence properties.

The natural development of this study is taking into consideration an entropy estimator which considers the continuous nature of returns and has the optimality properties.

### References


1. Bouzebda, S., Elhattab, I. **Uniform in bandwidth consistency of the kernel-type estimator of the Shannon's entropy**, Comptes Rendus Mathematique, Volume 348, Issues 5-6, March 2010, Pages 317-321, ISSN 1631-073X, DOI: 10.1016/j.crma.2009.12.007, (http://www.sciencedirect.com/science/article/B6X1B-4YG7P7T-1/2/87b80a60bc23cec17d06847c81598a7b).
2. Fix, E., Hodges, J. **Nonparametric discrimination: Consistency Properties**, Report 11, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.
3. Lorentz, Rudolph A. H. **On the entropy of a function**. J. Approx. Theory , 2009, 158, 2 (June 2009), pp. 145-150.
4. Nagelkerke, N. **A Note on a General Definition of the Coefficient of Determination**, Biometrika, vol. 78, no. 3, pp. 691–692, 1991.
5. Parzen, E. **On The Estimation Of Probability Density Function And Mode**, The Annals Of Mathematical Statistics, 33, pp. 1065-1076, 1962.
6. Risso, A. **The Informational Efficiency and the Financial Crashes**, Research in International Business and Finance, 2008, Vol. 22, pp. 396-408.
7. Sain, S. **Adaptive Kernel Density Estimation**, Ph.D. Thesis, Houston, Texas, 1994.
8. Saporta, G. **Probabilité, Analyse des Données et Statistique** , Ed.Technip,1990.
9. Scott, D. W. **Multivariate Density Estimation: Theory, Practice And Visualisation**, New York, John Wiley, 1992.

10. Silverman, B. W. **Density Estimation For Statistics And Data Analysis**, London, Chapman and Hall, 1986.

[1] **Acknowledgments**
This paper was co financed from the European Social Fund, through the Human Resources Development Operational Sectorial Program 2007-2013, project number POSDRU/89/1.5/S/59184, „Performance and excellence in the post-doctoral economic research in Romania".