Measuring the informational efficiency in the Stock Market

Wiston Adrián Risso
Department of Economics
University of Siena
risso@unisi.it

Outline

- Informational Efficiency & the Efficient Market Hypothesis (EMH)
- Some measures of efficiency
- Symbolic analysis & Shannon Entropy
- Examples: DotCom bubble (2000), Real Estate bubble (2007) and Banks (2009)
- Relationship with risk measures

Informational Efficiency & EMH

- The Efficient Market Hypothesis (EMH) in its weakly version, assumes that all information provided by the past prices is already embodied in the present prices.
- The most used and common framework for Stock prices has been the random walk

$$P(t) = P(t-1) + \varepsilon(t) \qquad \varepsilon(t) \neg\, i.i.d(0,\sigma)$$

- Due to the Efficiency of the market the stock prices are completely random and prediction is impossible.
- This Idea was supported by many scholar: Malkiel (1973), Fama (1965).
- Michael Jensen of Harvard University wrote in 1978 that "*the efficient market hypothesis is the best-established fact in all of social science.*"

Informational Efficiency & EMH

- According to Timmerman and Granger (2004) "*The behavior of market participants induce returns that obey the EMH, otherwise there would exist a "money-machine" producing unlimited wealth, which cannot occur in a stable economy.*"

- However, some "stylized facts" (fat tails, and volatility clustering) and critical events like the 1987 crisis made some scholars study the possibility of nonlinearities in the evolution of prices. See Hsieh (1991, 1995), Peters (1994, 1996), LeBaron (1994).

- **Question**: *If the market is not always informational efficient, can we find a measure of the level of informational efficiency? Is there a relationship between this measure and the financial risk*

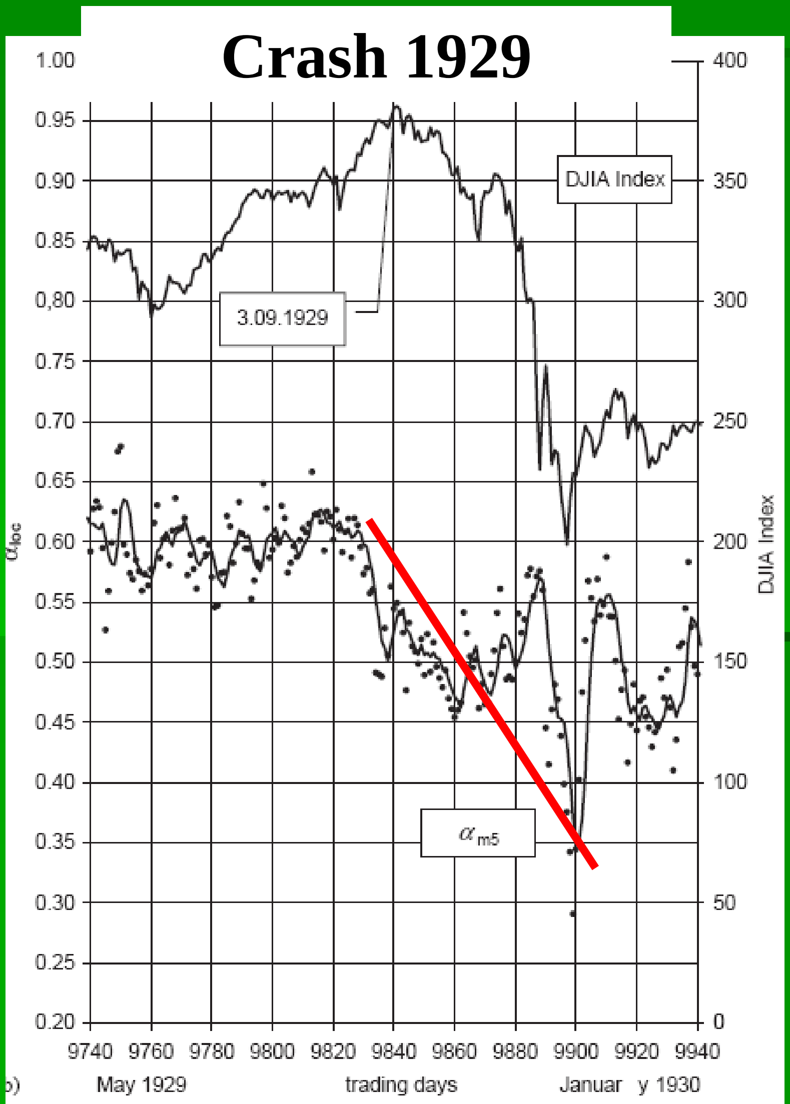
Some measures of efficiency

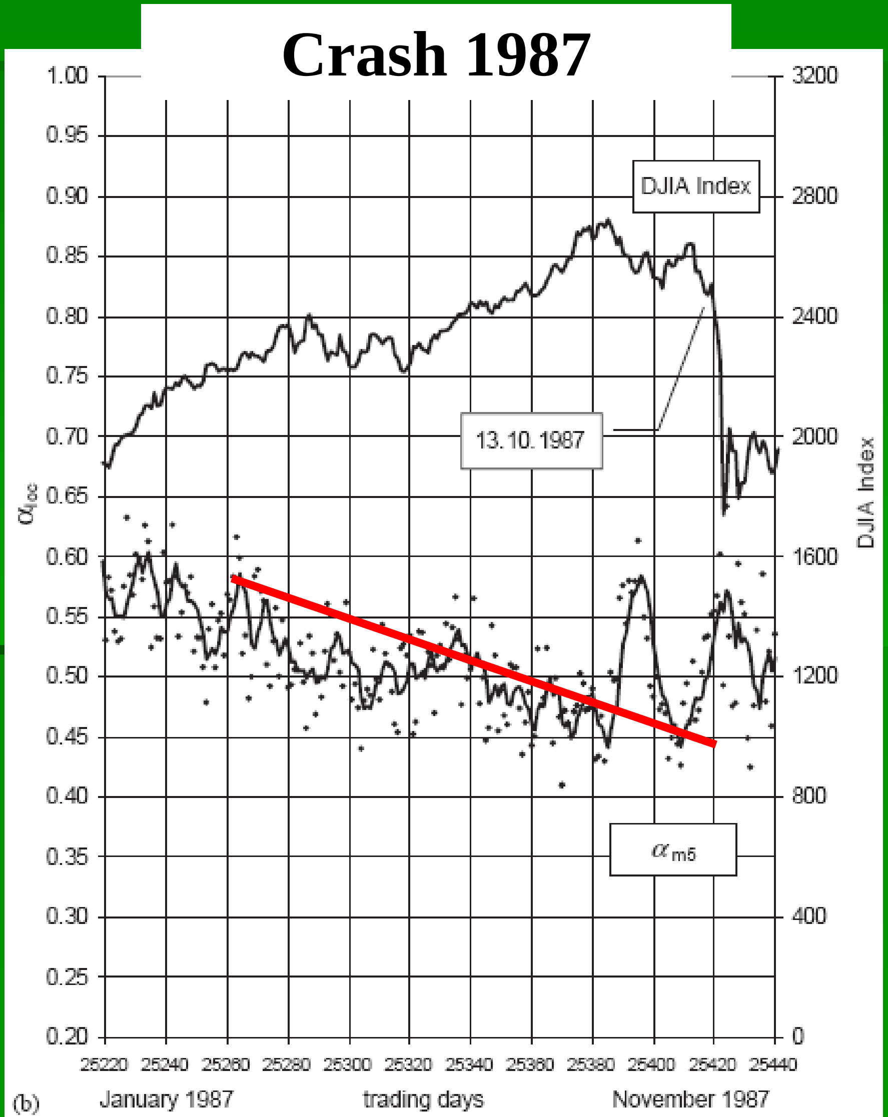
- Even if economists did not defined a measure of informational efficiency, Econophysics has proposed some measures.

- **HURST coefficient**: proposed by Hurst when studying the flood of river Nile. It was applied to the stock markets (see Peters 1994, 1996, Grech and Mazur 2004 and Cajueiro and Tabak 2004, 2005 among others)
- A Hurst exponent equal to 1/2 corresponds to a completely random process. Therefore an inefficient market should produce long memory and a Hurst >1/2
- **Criticism:** Bassler et. al (2006) and McCauley et al. (2007) assert that Hurst exponent different from ½ does not necessarily imply long time correlations.

Some measures of efficiency

- Taken from Grech and Mazur (2004), *Physica A*





Some measures of efficiency

- **Approximate entropy (ApEn)**: Pincus (1991) and Pincus and Singer (1996) proposed the *ApEn* to quantified the randomness in time series. When the time series data have a high degree of randomness, the *ApEn* is large. Oh, Kim and Eon (2007) use the measure in the financial markets with the embedding dimension $m=2$ and the distance $r=20\%$ of the standard deviation of the time series.

Symbolic analysis & Shannon Entropy

- I proposed a measure of efficiency in two steps: 1)Symbolization of the returns, 2) Shannon Entropy is applied to measure the quantity of information. (Risso, 2008) (Risso, 2009)
- **First**: Using the Symbolic Time Series Analysis (STSA) we can obtain richer information, transforming the data (Real) into a symbol time series of only few values (discrete). According to Daw et. al (2003) we can detect the very dynamic of the process when it is highly affected by noise. Ex.: r(t) are the stock returns.

$$\begin{cases} r(t) < 0 & s(t) = 0 \\ r(t) > 0 & s(t) = 1 \end{cases}$$

Symbolic analysis & Shannon Entropy

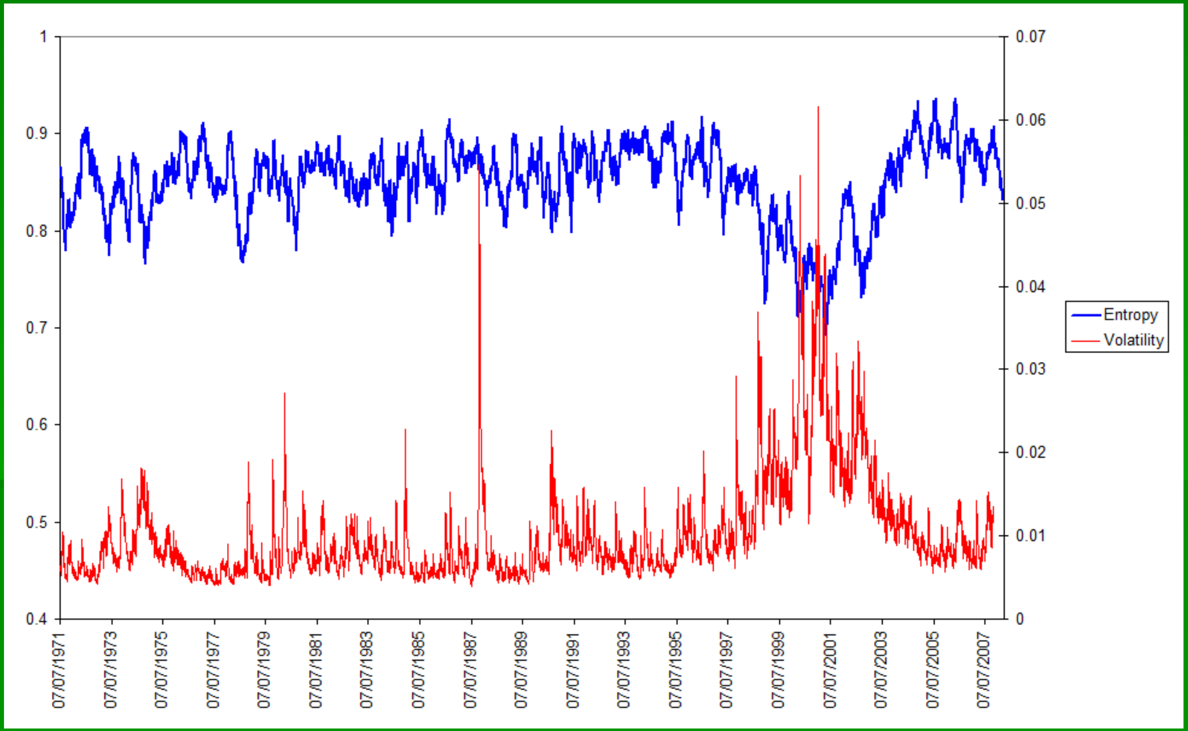
- **<u>Second</u>**: Normalized Shannon Entropy is applied to quantified the information in the series.

$$H = \left(\frac{-1}{\log_2(n)} \right) \sum p_i \log_2 p_i$$

- p_i is computed as the frequency of the event i appears in the series.
- Maximum efficiency when $H=1$, minimum when $H=0$

Ex.: DotCom Bubble 2000

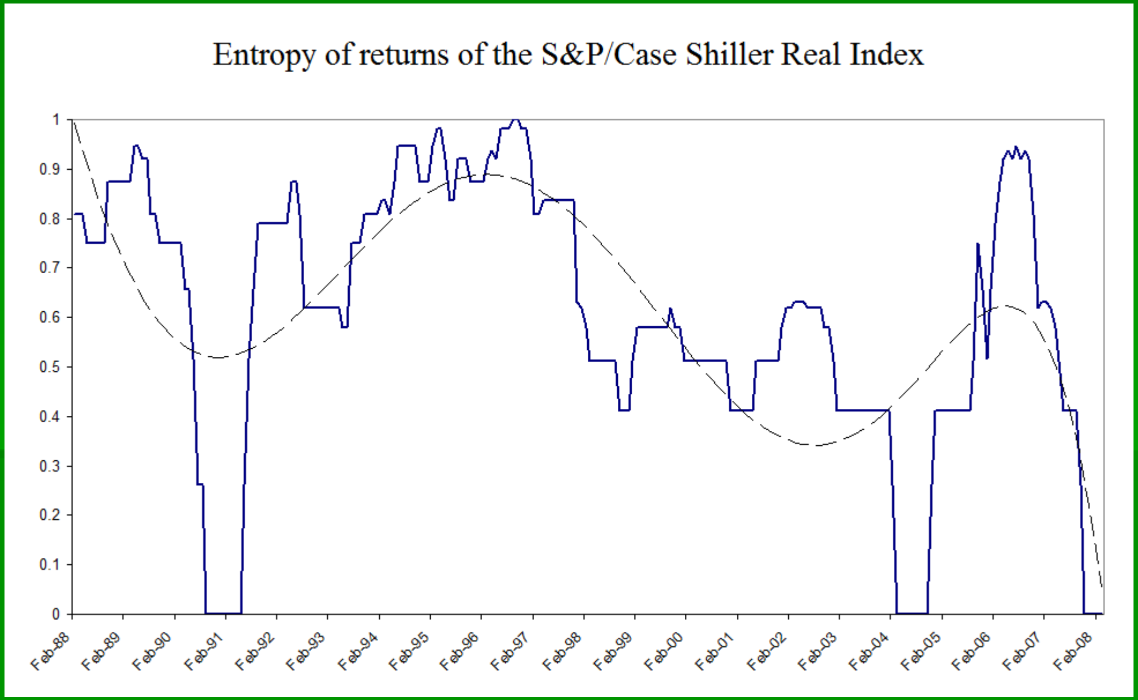
- **<u>DotCom Bubble:</u> The NASDAQ index. Crisis of 2000.**
- Daily data for the NASDAQ composite index from February 5, 1971 to April 3, 2008.



- *v=100*, sequence of 4 days. Clear cluster of inefficiency between August 17, 1998 and September 11, 2003 (minimum on April 27, 2001 H=0.693). Maximum Drawdown was 82.70% produced on October 9, 2002.

Ex.: Real Estate bubble 2007

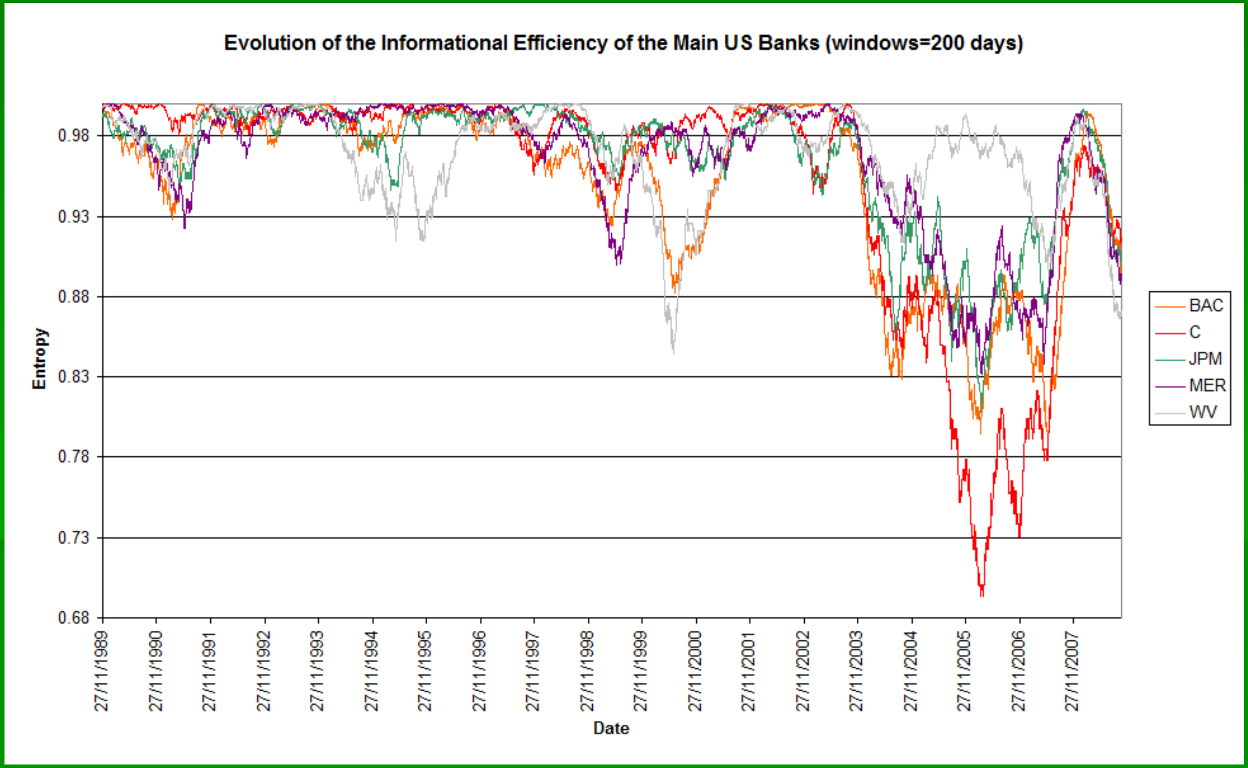
- **Real Estate Bubble**: Once the DotCom bubble burst, investors purchased real estate which many believed to be more reliable investment. We use the inflation adjusted S&P/Case Shiller Index in order to measure the real prices in the US housing market for the period January 1987 to March 2008.



- The periods 1990-1994 and 2005-March/2008 present negative real returns related with the periods of crisis in the sector.

Ex.: Banks (2009)

- **Banks**: There is a cluster of inefficiency for Bank of America (BAC), Citigroup (C), JP Morgan (JPM), Merrill Lynch (MER), Wachovia (WV) after 2003



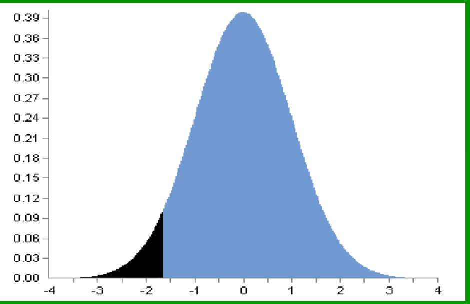
- The periods 1990-1994 and 2005-March/2008 present negative real returns related with the periods of crisis in the sector.

Relationship with risk measures

- **<u>Some Risk measures</u>**:
- *1)* ***<u>Volatility (Annualized Standard Deviation).</u>*** The standard deviation of return measures the average deviations of a return series from its mean, and is often used as a measure of risk. Note however that this definition includes in a symmetrical way both abnormal gains and abnormal losses.

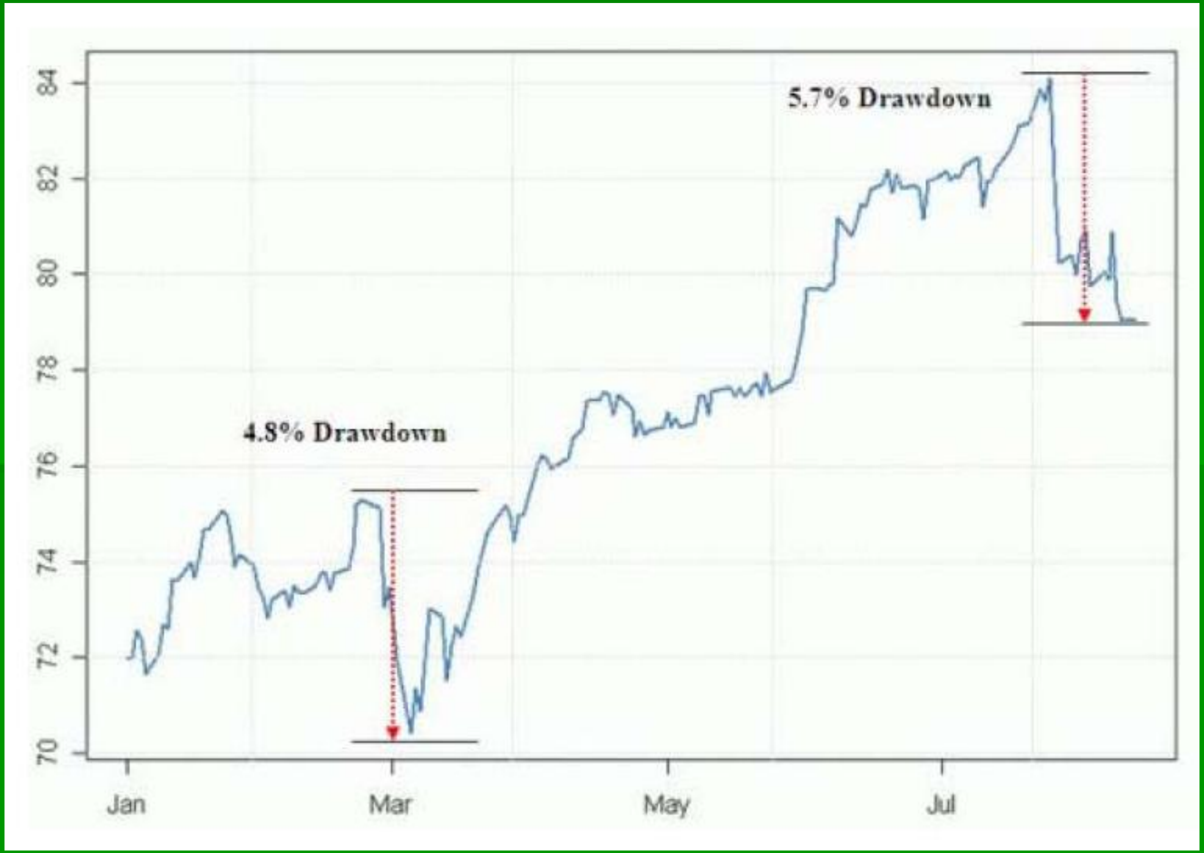
$$ASD = \sqrt{\frac{\sum_{t=1}^{T}(R_t)^2 - \frac{\left(\sum_{t=1}^{T} R_t\right)^2}{T}}{T}}\, x\sqrt{12}$$

- *2)* ***<u>Value at Risk (VaR).</u>*** A interesting notion is the probability of extreme losses, or, equivalently, the value-at-risk (VaR). The definition means that a loss equal or greater than the VaR over a time interval of *t=1* month (for example) happens 5 times over 100 months. Note, that this definition does not take into account the fact that losses can accumulate on consecutive time intervals *t*.



Relationship with risk measures

- *3) **<u>Maximum Drawdown</u>***. It is the largest percentage drop in your account between equity peaks. In other words, it's how much money you lose until you get back to breakeven.



- ..

Relationship with risk measures

According to Johnston and DiNardo (1997) the present logit model is given by:

$$p(y_i = 1) = \frac{\exp(\alpha + \beta H_i)}{1 - \exp(\alpha + \beta H_i)}$$

Logit Model: Econometric model where the dependent variable is the probability of one binary variable (ex.: Crash and no-crash).

Example: Real Estate Bubble

Observations = 242
LR $Chi^2(1) = 12.85$
Prob. > Chi^2 0.0003[a]
Log Likelihood = -46.353
Pseudo-$R^2 = 0.1758$

Crash prob.[c]	Coefficients	Standard error (σ)	t=coeff./σ	p-Value>|t|
Entropy (β)	-4.093	1.142	-3.58	0.000[b]
Constant (α)	-0.802	0.454	-1.77	0.077

The results were obtained with STATA program. Source: own calculations

[a] It indicates that the model is significant at 5%

[b] It indicates that the coefficients are significant at 5%

[c] It is the estimation of Eq. (3)

Relationship with risk measures

Example: DotCom Bubble

Table 1: Logit Model for the relationship between probability of crash and efficiency in the Nasdaq index

					No. Obs.=	9271
Crash Prob.[(a)]	**Coefficients**	**Standard error (σ)**	**t=Coeff./σ**	**p-value > |t|**	Log likelihood =	-426.99
Constant (α)	17.829	1.723	10.35	0.000	Wald chi^2(1)=	156.96
Entropy (β)	-27.284	2.178	-12.53	0.000	Prob > chi^2	0.000
					Pseudo-R^2=	0.1724
Crash Prob.[(b)]	**Coefficients**	**Standard error (σ)**	**t=Coeff./σ**	**p-value > |t|**	Log likelihood =	-638.99
Constant (α)	20.359	1.318	15.45	0.000	Wald chi^2(1)=	317.6
Entropy (β)	-29.716	1.667	-17.82	0.000	Prob > chi^2	0.000
					Pseudo-R^2=	0.2171

The results were obtained with STATA 9.0 program. Source: own calculations. v=100 days and sequence of 4 days produce the best fit

(a) Estimation of equation (4) using the definition of crash as losses larger than the mean minus 3 std. dev.

(b) Estimation of equation (4) using the second definition of crash, including high positive returns

Relationship with risk measures

- Relationship between probability of Crash and the entropy in 5 different markets (USA, Mexico, Malaysia, Japan and Russia)

Table 11
Logit model for five stock markets

$$p(y=1)=\frac{\exp(\alpha+\beta_0 H+\beta_1 \mathrm{MEX}+\beta_2 \mathrm{MAL}+\beta_3 \mathrm{JAP}+\beta_4 \mathrm{RUS})}{1-\exp(\alpha+\beta_0 H+\beta_1 \mathrm{MEX}+\beta_2 \mathrm{MAL}+\beta_3 \mathrm{JAP}+\beta_4 \mathrm{RUS})}$$

Observations = 11552
LR Chi^2 (5) = 235.27

Log likelihood = −801.75

Prob. > Chi^2	0.0000[a]
Pseudo-R^2 = 0.1280	

Crash prob.[c]	Coefficients	Standard error (σ)	t = coeff./σ	p-Value > |t|
Efficiency (β_0)	−55.07	8.56	−6.44	0.000[b]
MEX (β_1)	0.67	0.48	1.41	0.160
MAL (β_2)	1.01	0.46	2.18	0.029[b]
JAP (β_3)	0.79	0.50	1.58	0.113
RUS (β_4)	2.41	0.43	5.56	0.000[b]
Constant (α)	48.60	8.48	5.73	0.000[b]

The results where obtained with STATA program. *Source*: own calculations.

[a] It indicates that the model is significative at 5%.

[b] It indicates that the coefficients are significative at 5%.

[c] It is the estimation of Eq. (9).