

Misurare l'efficienza informativa dei mercati finanziari.

Lucio M. Calcagnile

Scuola Normale Superiore

2 febbraio 2011

Schema del seminario

- Efficienza informativa (mercato efficiente, *ipotesi dei mercati efficienti*)
- Alcuni tentativi in letteratura di misurarla (Giglio *et al.* (2008), Shmilovici *et al.* (2009), Risso (2009))
- Misurare l'efficienza con l'entropia di Shannon: studio su dati ad alta frequenza

Efficienza informativa

Un mercato *efficiente* è un mercato efficiente nel processare le informazioni. In un mercato efficiente, i prezzi dei beni osservati ad ogni istante sono basati su una corretta valutazione di tutta l'informazione disponibile.

«The modern literature in economics begins with Samuelson (1965), whose contribution is neatly summarized by the title of his article: “Proof that Properly Anticipated Prices Fluctuate Randomly”. In an informationally efficient market price changes must be unforecastable if they are properly anticipated, i.e., if they fully incorporate the expectations and information of all market participants.» Campbell, Lo, MacKinlay (1997)

Fama (1970): «A market in which prices always “fully reflect” available information is called “efficient”.»

Efficienza informativa

Jensen (1978): «A market is efficient with respect to information set θ_t if it is impossible to make economic profits by trading on the basis of information set θ_t .»

Malkiel (1992): «A capital market is said to be efficient if it fully and correctly reflects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set, ϕ , if security prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set, ϕ , implies that it is impossible to make economic profits by trading on the basis of ϕ .»

Ipotesi dei mercati efficienti

Ipotesi dei mercati efficienti (EMH):

- (forma debole) i prezzi riflettono completamente l'informazione contenuta nei prezzi passati
- (forma semi-forte) i prezzi riflettono tutta l'informazione rilevante pubblicamente disponibile
- (forma forte) i prezzi riflettono tutte le informazioni pubbliche e private

EMH: uno sguardo formale

Idealmente, un mercato efficiente è popolato da agenti razionali che assimilano rapidamente ogni informazione rilevante per la determinazione del prezzo e aggiustano i prezzi di conseguenza. Gli agenti processano efficientemente le informazioni e le incorporano immediatamente nei prezzi. Sotto l'EMH il prezzo P_t incorpora tutta l'informazione rilevante e il solo motivo perché esso cambi è l'arrivo di notizie o nuove informazioni fra i tempi t e $t + 1$.

Gli errori di previsione (*rendimenti in eccesso*)

$$\varepsilon_{t+1} = R_{t+1} - \mathbb{E}_t R_{t+1}$$

devono essere a media nulla e scorrelati da ogni informazione Θ_t disponibile al tempo t .

Verificabilità/falsificabilità dell'EMH

«The notion of market efficiency is not a well-posed and empirically refutable hypothesis. To make it operational, one must specify additional structure, e.g., investors' preferences, information structure, etc. But then a test of market efficiency becomes a test of several auxiliary hypotheses as well, and a rejection of such a joint hypothesis tells us little about which aspect of the joint hypothesis is inconsistent with the data.»

A. Lo (2000)

L'ipotesi della passeggiata aleatoria (RWH)

Esistono varie definizioni dell'ipotesi della passeggiata aleatoria. Esse stabiliscono condizioni che incorporano l'idea che i prezzi variano in maniera casuale e non predicibile.

$$\ln P_{t+1} = \mu + \ln P_t + \varepsilon_{t+1}$$

- 1 $\varepsilon_t \sim \text{IID}(0, \sigma^2)$
- 2 $\mathbb{E}[\varepsilon_t] = 0, \text{Cov}(\varepsilon_t, \varepsilon_{t+\tau}) = 0 \quad \forall t, \forall \tau > 0$

Misurare l'efficienza: Giglio et al., 2008

R. Giglio, R. Matsushita, S. Da Silva, *The relative efficiency of stockmarkets*, *Economics Bulletin* **7** (2008), 1–12.

R. Giglio, R. Matsushita, A. Figueiredo, I. Gleria, S. Da Silva, *Algorithmic Complexity Theory and the Relative Efficiency of Financial Markets*, *Europhysics Letters* **84** (2008), 48005–48010.

Seguendo Campbell, Lo, MacKinlay (1997): l'efficienza assoluta è un'idealizzazione. Non ha senso vedere se un mercato è efficiente o no. Sono sempre presenti inefficienze (→ informazione ridondante). Si dovrebbe cercare di quantificare quanto il mercato è lontano dall'efficienza ideale, con misure di lontananza dalla casualità. Usano il concetto di efficienza relativa, di un mercato rispetto a un altro.

Misurare l'efficienza: Giglio et al., 2008

La complessità algoritmica di Kolmogorov $C(s)$ di una stringa s è la lunghezza del più corto programma che la produce. Ma la funzione $s \mapsto C(s)$ non è calcolabile.

1976: Lempel e Ziv propongono una misura di complessità algoritmica (c) di una stringa.

1987: Kaspar e Schuster presentano un modo facile per calcolarla. Si ricostruisce la stringa data a partire dalla stringa vuota con una successione di operazioni che consistono nel copiare una sottostringa e aggiungere un simbolo. c è data dal numero di tali operazioni (+1 se l'ultima operazione è la sola copiatura di una sottostringa).

0000000000 $c = 2$

0101010101 $c = 3$

0110001001 $c = 5$

Misurare l'efficienza: Giglio *et al.*, 2008

La complessità algoritmica cresce con la lunghezza della stringa e per una stringa casuale $c \sim \frac{n}{\log_2 n} = r$.

Si calcola $LZ = \frac{c}{r}$ (misura di complessità rispetto alla stringa casuale).

Misurare l'efficienza: Giglio et al., 2008

- 36 indici di mercati azionari
- 37 titoli azionari di singole società del NYSE, NASDAQ, BOVESPA
- 20 cambi dollaro/altra moneta
- 7 anni di dati giornalieri (da luglio 2000 a luglio 2007), 2000 punti
- simbolizzazione dei rendimenti a tre simboli ($\{r_t \leq -b\}$, $\{|r_t| < b\}$, $\{r_t \geq b\}$, $b = 0,0025$)
- finestre mobili di ampiezza 1000, $LZ = \frac{c}{r}$ per ogni finestra
- efficienza relativa = $\frac{\text{n. di finestre } LZ > 1}{\text{totale}}$

Misurare l'efficienza: Giglio et al., 2008

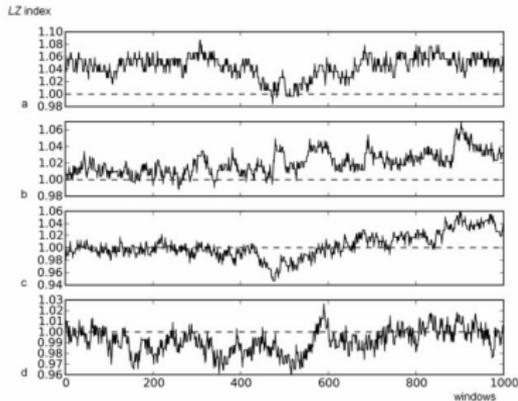


Figure 4. LZ index evolution over 1,000 sliding windows for (a) a computer-generated pseudo-random series (average LZ = 1.0180), (b) returns of the Dow Jones (average LZ = 1.0201), (c) returns of the Shanghai Composite (average LZ index = 1.0032), and (d) returns of the Karachi 100 (average LZ index = 0.9918).

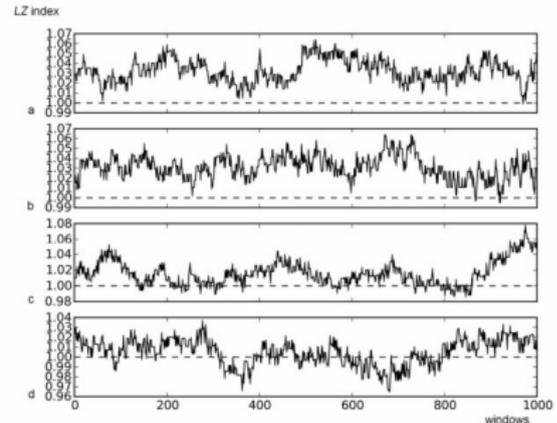


Figure 5. Evolution of the LZ index using 1,000 sliding windows for (a) Coca-Cola (100% efficient), (b) Yahoo (99.65% efficient), (c) Vale (92.75% efficient), and (d) Anacruz (66.67% efficient).

Table 1. The relative efficiency of selected stockmarket indices

Stock Exchange	Country	Average LZ Index	Degree of Efficiency*, %
S&P 500	USA	1.0232	99.1
DAX 30	GER	1.0257	98.4
Nikkei 225	JPN	1.0432	98.2
All Ordinaries	AUS	1.0246	97.8
ATX	AUT	1.0173	97.4
Dow Jones	USA	1.0201	95.4
Korea Composite	KOR	1.0163	94.9
Tel-Aviv 100	ISR	1.0187	92.9
Hang Seng	HKG	1.0151	91.5
Straits Times	SIN	1.0153	90.3
CAC 40	FRA	1.0138	88.4
Helsinki General	FIN	1.0149	88.4
Kuala Lumpur SE	MAS	1.0158	88
FTSE 100	UK	1.0106	86.6
Prague X	CZE	1.0139	81
Bel 20	BEL	1.0118	80.4
IBC	VEN	1.0110	79.9
Madrid General	ESP	1.0201	79.3
Swiss Market	SUI	1.0101	78.4
Nasdaq Composite	USA	1.0080	75.4
Amsterdam EX	NED	1.0100	74.4
Bovespa	BRA	1.0127	67.8
IPC	MEX	1.0060	64
Merval	ARG	1.0050	62.9
Jakarta Composite	IDN	1.0054	62.1
Istanbul 100	TUR	1.0085	61.3
Moscow Times	RUS	1.0050	59.2
Copenhagen	DEN	1.0025	58.7
Athex Composite	GRE	1.0048	56.9
Bombay SE	IND	1.0010	53.3
Taiwan Weighted	TPE	1.0006	50.3
Shanghai Composite	CHN	1.0032	49.5
Philippines	PHI	0.9987	43.1
Lima General	PER	0.9903	37.9
Karachi 100	PAK	0.9918	23.7
Colombo SE	SRI	0.9795	10.5

* Hits above the threshold LZ=1

Table 2. The relative efficiency of selected company stocks

Company	Stock Exchange	Average LZ Index	Degree of Efficiency*, %
Amazon	NYSE	1.0416	100
Coca-Cola	NYSE	1.0324	100
P&G	NYSE	1.0264	99.97
Intel	Nasdaq Composite	1.0292	99.92
eBay	Nasdaq Composite	1.0377	99.8
General Electric	NYSE	1.0274	99.66
Yahoo	Nasdaq Composite	1.0310	99.65
Texaco	NYSE	1.0264	99.46
Cisco	Nasdaq Composite	1.0357	99.44
Petrobras	Bovespa	1.0284	99.43
Pfizer	NYSE	1.0327	99.39
HP	NYSE	1.0298	99.38
Microsoft	Nasdaq Composite	1.0286	99.25
Goldman Sachs	NYSE	1.0311	98.78
J&J	NYSE	1.0275	98.73
Unilever	NYSE	1.0297	98.44
Nissan	Nasdaq Composite	1.0178	97.58
Merrill Lynch	NYSE	1.0279	97.33
JP Morgan	NYSE	1.0281	96.7
Oracle	Nasdaq Composite	1.0206	94.93
Citigroup	NYSE	1.0314	94.59
Vale	Bovespa	1.0193	92.75
Embraer	Bovespa	1.0258	91.59
Itau	Bovespa	1.0183	86.74
FedEx	NYSE	1.0186	86.6
Bradesco	Bovespa	1.0172	85.88
Exxon	NYSE	1.0161	85.56
Ford	NYSE	1.0152	84.26
Marcopolo	Bovespa	1.0072	77.36
Americanas	Bovespa	1.0136	76.68
Ipiranga	Bovespa	1.0111	76.55
Toyota	NYSE	1.0100	76.32
Wal-Mart	NYSE	1.0074	71.42
Ambev	Bovespa	1.0108	70.27
Aracruz	Bovespa	1.0048	66.67
Duratex	Bovespa	1.0048	65.17
Celese	Bovespa	1.0005	50.03

Table 2. The relative efficiency of selected US dollar exchange rates

Currency	Country	Average and Variance of the LZ Index	Relative Efficiency*, %	Average and Variance of the DFA Exponent	Occurrences Inside the Interval 0.5 ± 0.06 , %
Pound Sterling	UK	1.0223 (0.0001)	99.81	0.5272 (0.0003)	99
Swedish Krona	SWE	1.0236 (0.0002)	99.71	0.5247 (0.0004)	97
Norwegian Krone	NOR	1.0314 (0.0002)	99.60	0.5187 (0.0003)	100
Euro	Eurozone	1.0253 (0.0001)	99.45	0.5126 (0.0002)	100
New Zealand Dollar	NZL	1.0248 (0.0001)	99.20	0.4958 (0.0004)	99
Swiss Franc	SUI	1.0169 (0.0001)	99.12	0.5061 (0.0004)	100
Icelandic Krona	ISL	1.0184 (0.0001)	97.48	0.5463 (0.0004)	77
Mexican Peso	MEX	1.0254 (0.0001)	96.58	0.4875 (0.0003)	100
Danish Krone	DEN	1.0223 (0.0002)	94.10	0.5120 (0.0004)	99
Brazilian Real	BRA	1.0156 (0.0001)	92.60	0.5179 (0.0005)	95
Canadian Dollar	CAN	1.0219 (0.0002)	90.07	0.4837 (0.0001)	100
South African Rand	RSA	1.0177 (0.0002)	86.41	0.5115 (0.0005)	98
Japanese Yen	JPN	1.0153 (0.0002)	85.51	0.5157 (0.0001)	100
Singapore Dollar	SIN	1.0074 (0.0002)	66.48	0.5167 (0.0016)	77
Australian Dollar	AUS	1.004 (0.0001)	63.41	0.5074 (0.0005)	100
Indian Rupee	IND	0.9957 (0.0003)	43.54	0.5138 (0.0021)	81
Colombian Peso	COL	0.9913 (0.0002)	21.98	0.5580 (0.0004)	51
Taiwan New Dollar	TPE	0.9794 (0.0004)	21.17	0.5638 (0.0007)	57
Chinese Yuan	CHN	0.9265 (0.0039)	17.94	0.3655 (0.0505)	40
Sri Lanka Rupee	SLI	0.9687 (0.0006)	11.84	0.5449 (0.0025)	48

* Occurrences above the threshold $LZ = 1$
 Daily data from July 2000 to July 2007 (2,000 observations)
 Variances in brackets

Misurare l'efficienza: Shmilovici *et al.*, 2009

A. Shmilovici, Y. Kahiri, I. Ben-Gal, S. Hauser, *Measuring the Efficiency of the Intraday Forex Market with a Universal Data Compression Algorithm*, *Computational Economics* **33** (2009), 131–154.

Usano un *modello markoviano di ordine variabile (VOM model)* (Rissanen, 1983) per testare l'EMH (forma debole) per 12 coppie di valute.

Primo studio che collega compressione-predizione a dati empirici finanziari.

Misurare l'efficienza: Shmilovici *et al.*, 2009

VOM model

- Inizialmente sviluppato per compressione dati, gli autori ne usano una variante come modello di previsione di successioni simboliche.
- Rappresenta le sequenze statisticamente significative come un albero (similmente al *Context Tree* – Rissanen, 1983) e viene usato per predire un simbolo (quello che massimizza la probabilità, dato il contesto).

Misurare l'efficienza: Shmilovici *et al.*, 2009

- Anno 2002.
- 12 coppie di valute.
- Dati *tick by tick* (prezzi *bid*), aggregati a 1, 5, 10, 15, 20, 25, 30 minuti.
- Simbolizzazione a tre simboli: \nearrow , \searrow , $=$ ($< 0,0003$).
- Finestre mobili di ampiezza 50, 75, 100. Su ciascuna finestra si applica il modello, poi si media sulle finestre.
- 252 ($12 \cdot 7 \cdot 3$) esperimenti.

Misurare l'efficienza: Shmilovici *et al.*, 2009

Esperimenti e test statistici su comprimibilità, predicibilità e simulazione di due strategie di scambi:

- 1 automatica, che esegue una transazione se si prevede \nearrow o \searrow , senza soglia di probabilità per la predizione;
- 2 con soglia di probabilità ($s = 0,5, 0,7, 0,8$), che esegue una transazione se si prevede \nearrow o \searrow con probabilità maggiore di s .

Misurare l'efficienza: Shmilovici *et al.*, 2009

In 248 casi su 252 il modello ha rilevato alta comprimibilità, predicibilità sopra casualità (maggiormente per le serie ad alta frequenza).

Strategie di scambi: la strategia 1 genera troppe transazioni, che a causa dei costi fissi di transazione si traducono in perdita. La strategia 2 ha dato un guadagno, comunque molto basso, in soli 5 casi su 252.

Efficienza:

- (a) i prezzi incorporano tutta l'informazione significativa;
- (b) non esistono strategie di scambio profittevoli.

Gli esperimenti negano (a) e confermano (b).

Inefficienza teorica \longleftrightarrow efficienza pratica.

Misurare l'efficienza: Shmilovici *et al.*, 2009

In 248 casi su 252 il modello ha rilevato alta comprimibilità, predicibilità sopra casualità (maggiormente per le serie ad alta frequenza).

Strategie di scambi: la strategia 1 genera troppe transazioni, che a causa dei costi fissi di transazione si traducono in perdita. La strategia 2 ha dato un guadagno, comunque molto basso, in soli 5 casi su 252.

Efficienza:

- (a) i prezzi incorporano tutta l'informazione significativa;
- (b) non esistono strategie di scambio profittevoli.

Gli esperimenti negano (a) e confermano (b).

Inefficienza teorica \longleftrightarrow efficienza pratica.

Misurare l'efficienza: Risso, 2009

W. A. Risso, *The Informational Efficiency: The Emerging Markets Versus the Developed Markets*, Applied Economics Letters **16** (2009), 485–487.

- scopo: studiare le differenze dei livelli di efficienza informativa fra i mercati emergenti e quelli sviluppati
- analisi simbolica delle serie temporali ed entropia di Shannon come misura d'efficienza
- 20 indici di mercati azionari
- 10 anni di dati giornalieri (da luglio 1997 a dicembre 2007)
- simbolizzazione dei rendimenti a due simboli ($\{r_t < \bar{r}\}$, $\{r_t > \bar{r}\}$)
- “I risultati sembrano confortare l'ipotesi che i mercati azionari emergenti (in particolare quelli dei paesi ex socialisti) sono piú inefficienti rispetto a quelli sviluppati.”

Misurare l'efficienza: Risso, 2009

Table I: Ranking of Informational Efficiency for different stock indices

Rank	Index	Country	Entropy	Rank	Index	Country	Entropy
1	TSEC	Taiwan	0.9833	11	MERVAL	Argentina	0.9791
2	NIKKEI	Japan	0.9806	12	ATX	Austria	0.9791
3	Straits Times	Singapore	0.9805	13	JKSE	Indonesia	0.9787
4	IPC	Mexico	0.9804	14	SSMI	Switzerland	0.9783
5	TASE	Israel	0.9801	15	Hang Seng	Hong Kong	0.9777
6	DJIA	USA	0.9800	16	KLSE	Malaysia	0.9769
7	AEX	Holland	0.9799	17	PX GLOB	Czech. Rep.	0.9768
8	KOSPI	South Korea	0.9795	18	RTS	Russia	0.9751
9	DAX	Germany	0.9794	19	CMA	Egypt	0.9669
10	FTSE	UK	0.9794	20	SBI 20	Slovenia	0.9481

Based on the own calculations.

Dati ad alta frequenza

Exchange Traded Funds dell'American Stock Exchange.

simbolo	ETF: nome	da	a	righe		buchi		
				a	b	c	b/(b+a)	c/(c+a)
DIA	Diamonds	02/01/03	12/01/09	692344	20866	9923	0,0293	0,0141
QQQQ	PowerShares QQQ	12/01/04	12/01/09	497295	11192	2354	0,0220	0,0047
SPY	Standard & Poors 500	02/01/03	12/01/09	710424	8725	2473	0,0121	0,0035
SHY	iShares Lehman 1-3 Year Treasury	02/01/03	12/01/09	260195	305929	119422	0,5404	0,3146
TLT	iShares Lehman 20+ Year Treasury	02/01/03	12/01/09	348256	223126	96275	0,3905	0,2166
IEF	iShares Lehman 7-10 Year Treasury	02/01/03	12/01/09	196136	369355	108412	0,6532	0,3560
EEM	iShares MSCI Emerging Markets	04/11/03	12/01/09	521801	147213	45815	0,2200	0,0807
EWA	iShares MSCI-Australia	02/01/03	12/01/09	343841	332948	97924	0,4920	0,2217
EWO	iShares MSCI-Austria	02/01/03	12/01/09	111276	410435	68543	0,7867	0,3812
EWK	iShares MSCI-Belgium	02/01/03	12/01/09	76540	516600	53576	0,8710	0,4118
EWZ	iShares MSCI-Brazil	02/01/03	12/01/09	493804	201134	50882	0,2894	0,0934
EWG	iShares MSCI-Canada	02/01/03	12/01/09	224553	329705	82278	0,5949	0,2682
EFA	iShares MSCI-EAFE	02/01/03	12/01/09	585122	108544	52059	0,1565	0,0817
EZU	iShares MSCI-EMU	08/21/07	12/01/09	49126	171058	31571	0,7769	0,3912
EWQ	iShares MSCI-France	02/01/03	12/01/09	84546	537163	54874	0,8640	0,3936
EWG	iShares MSCI-Germany	02/01/03	12/01/09	275973	397078	107097	0,5900	0,2796
EWH	iShares MSCI-Hong Kong	02/01/03	12/01/09	412744	268206	103656	0,3939	0,2007
EWI	iShares MSCI-Italy	02/01/03	12/01/09	70526	541776	47894	0,8848	0,4044
EWJ	iShares MSCI-Japan	02/01/03	12/01/09	612298	83106	47721	0,1195	0,0723
EWM	iShares MSCI-Malaysia	02/01/03	12/01/09	302214	369739	101578	0,5502	0,2516
EWV	iShares MSCI-Mexico	02/01/03	12/01/09	410213	258874	58720	0,3869	0,1252
EWN	iShares MSCI-Netherlands	02/01/03	12/01/09	31535	236496	22034	0,8823	0,4113
EPP	iShares MSCI-Pacific-Ex Japan	02/16/07	12/01/09	142932	130135	51567	0,4766	0,2651
EWS	iShares MSCI-Singapore	02/01/03	12/01/09	335994	338642	91614	0,5020	0,2142
EZA	iShares MSCI-South Africa	02/16/07	12/01/09	74239	195955	42477	0,7252	0,3639
EWY	iShares MSCI-South Korea	02/01/03	12/01/09	423962	254565	80665	0,3752	0,1599
EWV	iShares MSCI-Spain	02/01/03	12/01/09	61846	220942	36143	0,7813	0,3688
EWD	iShares MSCI-Sweden	02/01/03	12/01/09	121254	510481	72393	0,8081	0,3738
EWL	iShares MSCI-Switzerland	02/01/03	12/01/09	95800	545727	65478	0,8507	0,4060
EWT	iShares MSCI-Taiwan	02/01/03	12/01/09	438576	243972	79368	0,3574	0,1532
EWU	iShares MSCI-UK	11/30/05	12/01/09	145831	243400	73378	0,6253	0,3347
NM	iShares Russell 2000	02/01/03	12/01/09	675729	34896	18160	0,0491	0,0262
UH	iShares S&P MidCap 400	11/30/05	12/01/09	245497	146421	76676	0,3736	0,2380
UR	iShares S&P SmallCap 600	02/01/03	12/01/09	452851	231882	105734	0,3386	0,1893

Serie ad alta frequenza

10/07/2005, 11:53, 30.80, 30.83, 30.80	30.83	1300
10/07/2005, 11:54, 30.82, 30.84, 30.82	30.83	2500
10/07/2005, 11:55, 30.83, 30.86, 30.83	30.83	2600
10/07/2005, 11:56, 30.86, 30.86, 30.86	30.86	700
10/07/2005, 11:57, 30.86, 30.94, 30.86	30.94	11200
10/07/2005, 11:58, 30.95, 30.95, 30.94	30.95	3400
10/07/2005, 11:59, 30.97, 30.99, 30.95	30.98	7700
10/07/2005, 12:00, 30.97, 30.97, 30.97	30.97	4300
10/07/2005, 12:01, 30.98, 30.98, 30.98	30.98	100
10/07/2005, 12:02, 30.96, 30.98, 30.95	30.98	1500
10/07/2005, 12:04, 30.98, 31.00, 30.95	30.99	4600
10/07/2005, 12:05, 30.99, 30.99, 30.90	30.99	4900
10/07/2005, 12:06, 30.99, 30.99, 30.95	30.95	600
10/07/2005, 12:07, 30.99, 30.99, 30.99	30.99	200
10/07/2005, 12:08, 30.99, 30.99, 30.99	30.99	100
10/07/2005, 12:09, 30.99, 30.99, 30.97	30.99	3100
10/07/2005, 12:10, 30.99, 30.99, 30.99	30.99	800
10/07/2005, 12:11, 30.99, 31.02, 30.99	31.02	18800
10/07/2005, 12:12, 31.02, 31.03, 31.02	31.02	39000
10/07/2005, 12:13, 31.03, 31.03, 31.02	31.02	2100
10/07/2005, 12:15, 31.00, 31.03, 30.99	31.03	12800
10/07/2005, 12:16, 31.05, 31.05, 31.02	31.02	20300
10/07/2005, 12:18, 31.05, 31.05, 31.05	31.05	3600
10/07/2005, 12:19, 31.05, 31.06, 31.05	31.06	13200
10/07/2005, 12:20, 31.06, 31.06, 31.05	31.05	11000
10/07/2005, 12:21, 31.06, 31.06, 31.02	31.05	2800
10/07/2005, 12:22, 31.07, 31.07, 31.07	31.07	100
10/07/2005, 12:23, 31.07, 31.07, 31.07	31.07	18200
10/07/2005, 12:24, 31.07, 31.07, 31.06	31.07	7300
10/07/2005, 12:25, 31.07, 31.07, 31.07	31.07	1100
10/07/2005, 12:27, 31.07, 31.07, 31.07	31.07	1300
10/07/2005, 12:28, 31.07, 31.07, 31.07	31.07	3100

$$P_t$$

↓

$$r_t = \ln P_t - \ln P_{t-1}$$

↓

serie simbolica

Sorgente d'informazione ed entropia: uno sguardo informale

Sorgente d'informazione stazionaria a stati finiti:

- dispositivo che emette stringhe di simboli $x_0x_1x_2\dots$
- ogni x_i appartiene a un *alfabeto* finito A
- la probabilità di ricevere una data stringa finita non varia nel tempo

Entropia: misura dell'incertezza associata a una variabile aleatoria.
Informazione a bassa entropia:

- 0000100000001000010000000010100000010000000000100...
- 10110011011000101100001011001101100111...
- 101100 1 101100 0 101100 00 101100 1 101100 111...

Sorgente d'informazione ed entropia: uno sguardo informale

Sorgente d'informazione stazionaria a stati finiti:

- dispositivo che emette stringhe di simboli $x_0x_1x_2\dots$
- ogni x_i appartiene a un *alfabeto* finito A
- la probabilità di ricevere una data stringa finita non varia nel tempo

Entropia: misura dell'incertezza associata a una variabile aleatoria.
Informazione a bassa entropia:

- 0000100000001000010000000010100000010000000000100...
- 10110011011000101100001011001101100111...
- **101100 1 101100 0 101100 00 101100 1 101100 111...**

Sorgente d'informazione

Formalmente, una sorgente è un sistema dinamico $(\Omega, \mathcal{F}, \mu, \sigma)$:

- $\Omega = A^{\mathbb{N}}$
- \mathcal{F} è la σ -algebra generata dai cilindri $\{\omega_i = a_i\}$, $a_i \in A$
- μ è una misura di probabilità definita sugli eventi $E \in \mathcal{F}$
- $\sigma : \Omega \rightarrow \Omega$ è l'applicazione di traslazione (*shift*)

$$\sigma(a_0 a_1 a_2 a_3 \dots) = a_1 a_2 a_3 a_4 \dots$$

- stazionarietà $\leftrightarrow \mu$ σ -invariante
- è ben definita la probabilità di una stringa finita:
 $\mu(a_1 a_2 \dots a_k)$, $a_i^k \in A^k$

Sorgente d'informazione

Supporremo che le sorgenti siano *ergodiche*: per quasi ogni $\omega \in \Omega$ e per ogni stringa finita s ,

$$\lim_{n \rightarrow \infty} \frac{\#\{s \subseteq \omega_1 \dots \omega_n\}}{n - |s| + 1} = \mu(s)$$

Entropia di Shannon di una sorgente

$$H_k = - \sum_{\text{length}(\omega)=k} \mu(\omega) \log \mu(\omega)$$

$$h = \lim_{k \rightarrow \infty} \frac{H_k}{k}$$

- h rappresenta la quantità media di informazione che la sorgente trasmette con ogni simbolo emesso
- h misura il grado di incertezza che un lettore della sorgente ha circa il simbolo successivo che verrà emesso

Stimatori dell'entropia

$$h = \lim_{k \rightarrow \infty} -\frac{1}{k} \sum_{\text{length}(\omega)=k} \mu(\omega) \ln \mu(\omega)$$

$$H = -\sum \mu(\omega) \ln \mu(\omega) = \ln N - \frac{1}{N} \sum (\mu(\omega)N) \ln(\mu(\omega)N)$$

$$\hat{H}_{\text{naïve}} = \ln N - \frac{1}{N} \sum_i n_i \ln n_i$$

- Lo *stimatore* $\hat{H}_{\text{naïve}}$ è distorto:

$$\mathbb{E}[\hat{H}_{\text{naïve}}] - H < 0$$

Stimatori dell'entropia

$$\hat{H}_{\text{naïve}} = \ln N - \frac{1}{N} \sum_i n_i \ln n_i \quad \rightarrow \quad \hat{H}_G = \ln N - \frac{1}{N} \sum_i n_i G_{n_i}$$

- $G_{2n+1} = G_{2n} = -\gamma - \ln 2 + \frac{2}{1} + \frac{2}{3} + \frac{2}{5} + \dots + \frac{2}{2n-1}$
 $\gamma = 0.577215 \dots$ costante di Eulero
- $0 < H - \mathbb{E}[\hat{H}_G] < 0,1407 \dots \times \frac{M}{N}$

Simbolizzazione dei rendimenti

- due simboli: $\{r_t < 0\}$, $\{r_t > 0\}$
- tre simboli:
 - 1 $\{r_t < 0\}$, $\{r_t = 0\}$, $\{r_t > 0\}$
 - 2 $\{r_t < -b\}$, $\{-b \leq r_t \leq b\}$, $\{r_t > b\}$, $b > 0$
 - 3 $\{r_t < t_1\}$, $\{t_1 \leq r_t \leq t_2\}$, $\{r_t > t_2\}$, t_i i -esimo terzile

Simbolizzazione a due simboli

	simboli	n. "0"	n. "1"	% "0"	% "1"	stima h
IWM	558789	278022	280767	49,75	50,25	0,998716
IJH	224179	111591	112588	49,78	50,22	0,998464
EZU	44360	22192	22168	50,03	49,97	0,998233
SPY	625121	311778	313343	49,87	50,13	0,998094
EWZ	412507	205500	207007	49,82	50,18	0,997281
EZA	66917	33207	33710	49,62	50,38	0,997246
DIA	604522	301401	303121	49,86	50,14	0,997169
EPP	124410	61832	62578	49,70	50,30	0,996959
QQQQ	423305	211574	211731	49,98	50,02	0,996777
IJR	383077	190638	192439	49,76	50,24	0,996069
EEM	426480	211914	214566	49,69	50,31	0,995817
EWV	324522	162580	161942	50,10	49,90	0,995306
TLT	276797	137417	139380	49,65	50,35	0,994497
EWP	53099	26394	26705	49,71	50,29	0,992129
EWY	323584	161864	161720	50,02	49,98	0,990385
EWN	25186	12623	12563	50,12	49,88	0,990240
EWQ	66554	33104	33450	49,74	50,26	0,988696
EFA	446526	221737	224789	49,66	50,34	0,988648
EWI	54040	26623	27417	49,27	50,73	0,986155
EWC	169473	84203	85270	49,69	50,31	0,985423
EWG	192191	95343	96848	49,61	50,39	0,984525
EWU	107661	53704	53957	49,88	50,12	0,981817
EWD	93120	46321	46799	49,74	50,26	0,980997
EWA	240385	119822	120563	49,85	50,15	0,978246
EWL	70912	35504	35408	50,07	49,93	0,971652
IEF	153524	76300	77224	49,70	50,30	0,967818
EWK	55117	27344	27773	49,61	50,39	0,961603
EWO	81484	40383	41101	49,56	50,44	0,958074
EWT	255497	127946	127551	50,08	49,92	0,957215
EWB	256848	128430	128418	50,00	50,00	0,956197
EWS	192456	96188	96268	49,98	50,02	0,942466
EWM	160897	80169	80728	49,83	50,17	0,923845
EWJ	320941	160301	160640	49,95	50,05	0,860182
SHY	175662	87390	88272	49,75	50,25	0,855905

Simbolizzazione a tre simboli 1

	simboli	n. "0"	n. "1"	n. "2"	% "0"	% "1"	% "2"	stima h
IJH	245496	111591	21317	112588	45,46	8,68	45,86	0,997763
SPY	710423	311778	85302	313343	43,89	12,01	44,11	0,996251
DIA	692343	301401	87821	303121	43,53	12,68	43,78	0,996015
QQQQ	497294	211574	73989	211731	42,55	14,88	42,58	0,994624
EZU	49125	22192	4765	22168	45,17	9,70	45,13	0,994068
IJR	452850	190638	69773	192439	42,10	15,41	42,50	0,993258
IWM	675728	278022	116939	280767	41,14	17,31	41,55	0,992575
EPP	142931	61832	18521	62578	43,26	12,96	43,78	0,991986
EZA	74238	33207	7321	33710	44,73	9,86	45,41	0,991587
TLT	348255	137417	71458	139380	39,46	20,52	40,02	0,989113
EWV	410212	162580	85690	161942	39,63	20,89	39,48	0,988958
EWP	61845	26394	8746	26705	42,68	14,14	43,18	0,984078
EWZ	493803	205500	81296	207007	41,62	16,46	41,92	0,980895
EWC	224552	84203	55079	85270	37,50	24,53	37,97	0,980531
EWQ	84545	33104	17991	33450	39,16	21,28	39,56	0,979612
EWG	275972	95343	83781	96848	34,55	30,36	35,09	0,979338
EWU	145830	53704	38169	53957	36,83	26,17	37,00	0,979126
EWD	121253	46321	28133	46799	38,20	23,20	38,60	0,978808
EEM	521800	211914	95320	214566	40,61	18,27	41,12	0,978293
EFA	585121	221737	138595	224789	37,90	23,69	38,42	0,977291
EWI	70525	26623	16485	27417	37,75	23,37	38,88	0,976628
EWY	423961	161864	100377	161720	38,18	23,68	38,15	0,975687
EWN	31534	12623	6348	12563	40,03	20,13	39,84	0,975042
EWA	343840	119822	103455	120563	34,85	30,09	35,06	0,972783
IEF	196135	76300	42611	77224	38,90	21,73	39,37	0,972376
EWT	438575	127946	183078	127551	29,17	41,74	29,08	0,970034
EWL	95799	35504	24887	35408	37,06	25,98	36,96	0,968674
EWH	412743	128430	155895	128418	31,12	37,77	31,11	0,967164
EWK	76539	27344	21422	27773	35,73	27,99	36,29	0,962745
EWS	335993	96188	143537	96268	28,63	42,72	28,65	0,960896
EWO	111275	40383	29791	41101	36,29	26,77	36,94	0,954748
EWV	302213	80169	141316	80728	26,53	46,76	26,71	0,950251
EWJ	612297	160301	291356	160640	26,18	47,58	26,24	0,940276
SHY	260194	87390	84532	88272	33,59	32,49	33,93	0,908900

Simbolizzazione a tre simboli 2

	simboli	n. "0"	n. "1"	n. "2"	% "0"	% "1"	% "2"	stima h
IJH	245496	106135	32345	107016	43,23	13,18	43,59	0,996642
EZU	49125	20897	7294	20934	42,54	14,85	42,61	0,995127
QQQ	497294	193618	109737	193939	38,93	22,07	39,00	0,993992
IJR	452850	188410	74252	190188	41,61	16,40	42,00	0,99369
IWM	675728	276061	120806	278861	40,85	17,88	41,27	0,993265
EPP	142931	57578	27168	58185	40,28	19,01	40,71	0,991583
TLT	348255	128650	89325	130280	36,94	25,65	37,41	0,991256
EZA	74238	31691	10308	32239	42,69	13,89	43,43	0,99062
EWW	410212	162267	86306	161639	39,56	21,04	39,40	0,989726
EWP	61845	26001	9532	26312	42,04	15,41	42,55	0,98608
EWZ	493803	203651	84897	205255	41,24	17,19	41,57	0,983298
EWC	224552	77765	67997	78790	34,63	30,28	35,09	0,98158
EEM	521800	194850	129665	197285	37,34	24,85	37,81	0,981453
EFA	585121	217171	147660	220290	37,12	25,24	37,65	0,981017
EWG	275972	93648	87052	95272	33,93	31,54	34,52	0,981012
EWQ	84545	32857	18496	33192	38,86	21,88	39,26	0,980795
EWD	121253	45890	29009	46354	37,85	23,92	38,23	0,980353
EWJ	145830	51406	42839	51585	35,25	29,38	35,37	0,979953
EWI	70525	26376	16979	27170	37,40	24,08	38,53	0,979232
EWN	31534	12465	6655	12414	39,53	21,10	39,37	0,977385
EWY	423961	160172	103826	159963	37,78	24,49	37,73	0,977277
EWA	343840	116121	110751	116968	33,77	32,21	34,02	0,976066
IEF	196135	69560	56191	70384	35,47	28,65	35,89	0,973522
SPY	710423	244451	220379	245593	34,41	31,02	34,57	0,973295
EWT	438575	124731	189468	124376	28,44	43,20	28,36	0,970951
EWL	95799	35068	25760	34971	36,61	26,89	36,50	0,970932
EWH	412743	124555	163526	124662	30,18	39,62	30,20	0,968598
DIA	692343	236634	217626	238083	34,18	31,43	34,39	0,966957
EWS	335993	93441	149097	93455	27,81	44,38	27,81	0,963768
EWK	76539	27064	21996	27479	35,36	28,74	35,90	0,958638
EWO	111275	37969	34742	38564	34,12	31,22	34,66	0,955414
EWM	302213	77622	146337	78254	25,68	48,42	25,89	0,953887
EWJ	612297	156680	298640	156977	25,59	48,77	25,64	0,942602
SHY	260194	64090	131680	64424	24,63	50,61	24,76	0,918963

Simbolizzazione a tre simboli 3

	simboli	n. "0"	n. "1"	n. "2"	% "0"	% "1"	% "2"	stima h
IJH	245496	81835	81827	81834	33,33	33,33	33,33	0,9883
EZA	74238	24746	24746	24746	33,33	33,33	33,33	0,9866
TLT	348255	116085	116055	116115	33,33	33,32	33,34	0,9860
EZU	49125	16375	16375	16375	33,33	33,33	33,33	0,9839
IJR	452850	150958	150929	150963	33,34	33,33	33,34	0,9834
EWZ	493803	164610	164591	164602	33,34	33,33	33,33	0,9829
EWG	224552	74851	74850	74851	33,33	33,33	33,33	0,9824
EWG	275972	91999	91946	92027	33,34	33,32	33,35	0,9822
SPY	710423	236809	236801	236813	33,33	33,33	33,33	0,9821
EWU	145830	48611	48599	48620	33,33	33,33	33,34	0,9802
EWA	343840	114613	114614	114613	33,33	33,33	33,33	0,9788
QQQQ	497294	165771	165757	165766	33,33	33,33	33,33	0,9785
QQQQ	410212	136738	136733	136741	33,33	33,33	33,33	0,9780
EWP	61845	20615	20615	20615	33,33	33,33	33,33	0,9773
IWM	675728	225253	225223	225252	33,33	33,33	33,33	0,9767
EEM	521800	173936	173931	173933	33,33	33,33	33,33	0,9750
EWY	423961	141330	141306	141325	33,34	33,33	33,33	0,9732
DIA	692343	230781	230787	230775	33,33	33,33	33,33	0,9720
EFA	585121	195054	195023	195044	33,34	33,33	33,33	0,9712
EPP	142931	47650	47637	47644	33,34	33,33	33,33	0,9702
EWT	438575	127946	183078	127551	29,17	41,74	29,08	0,9700
IEF	196135	65378	65379	65378	33,33	33,33	33,33	0,9697
EWG	121253	40420	40409	40424	33,34	33,33	33,34	0,9696
EWG	412743	128430	155895	128418	31,12	37,77	31,11	0,9691
EWQ	84545	28183	28174	28188	33,33	33,32	33,34	0,9680
EWI	70525	23514	23501	23510	33,34	33,32	33,34	0,9671
EWN	31534	10514	10508	10512	33,34	33,32	33,34	0,9617
EWL	95799	31948	31915	31936	33,35	33,31	33,34	0,9612
EWK	76539	25513	25507	25519	33,33	33,33	33,34	0,9609
EWS	335993	96188	143537	96268	28,63	42,72	28,65	0,9609
EWO	111275	37093	37089	37093	33,33	33,33	33,33	0,9606
EWM	302213	80169	141316	80728	26,53	46,76	26,71	0,9503
EWJ	612297	160301	291356	160640	26,18	47,58	26,24	0,9403
SHY	260194	86736	86724	86734	33,34	33,33	33,33	0,9167

Serie temporali ad alta frequenza: vantaggi e svantaggi

Vantaggi

- piú informazioni sull'evoluzione dei prezzi
- statistica piú ricca

Svantaggi

- andamento intragiornaliero
- effetti di *microstruttura* del mercato non trascurabili (tanto piú marcati quanto piú alta è la frequenza dei dati): dipendono dai costi fissi di transazione, da come viene determinato il prezzo e in generale dalle regole del mercato
 - *bid-ask spread*
 - discretezza del prezzo

Un modello di microstruttura

Modello di Heston (1993) per l'evoluzione del prezzo vero

$$dp^*(t) = \left(\mu - \frac{v(t)}{2}\right)dt + \sigma(t)dB(t)$$
$$dv(t) = k(\alpha - v(t))dt + \gamma\sqrt{v(t)}dW(t).$$

Modello di Hasbrouck (1999) per il prezzo osservato

$$B_t = \Delta \left[\frac{P_t^*}{\Delta} - 1 \right], \quad A_t = \Delta \left[\frac{P_t^*}{\Delta} + 1 \right],$$

con $\Delta = \text{costo } bid = \text{costo } ask = \text{tick size}$, $P_t^* = e^{P_t^*}$.

$$P_t = B_t q_t + A_t (1 - q_t) \quad (\text{prezzo osservato})$$

con $q_t \sim \text{Bernoulli}(\frac{1}{2})$.

-  R. Giglio, R. Matsushita, S. Da Silva, *The relative efficiency of stockmarkets*, *Economics Bulletin* **7** (2008), 1–12.
-  R. Giglio, R. Matsushita, A. Figueiredo, I. Gleria, S. Da Silva, *Algorithmic Complexity Theory and the Relative Efficiency of Financial Markets*, *Europhysics Letters* **84** (2008), 48005–48010.
-  W. A. Risso, *The Informational Efficiency: The Emerging Markets Versus the Developed Markets*, *Applied Economics Letters* **16** (2009), 485–487.
-  A. Shmilovici, Y. Kahiri, I. Ben-Gal, S. Hauser, *Measuring the Efficiency of the Intraday Forex Market with a Universal Data Compression Algorithm*, *Computational Economics* **33** (2009), 131–154.