On Predictability and Profitability: Would AI induced Trading Rules be Sensitive to the Entropy of time Series

Nicolas NAVET – INRIA
France
nnavet@loria.fr

Shu-Heng CHEN – AIECON/NCCU
Taiwan
chchen@nccu.edu.tw

09/04/2008

Outline

- **Entropy Rate :** uncertainty remaining in the next information produced given knowledge of the past ➔ measure of predictability

- Questions :
  - Do stocks exhibit differing entropy rates?
  - Does low entropy imply profitability of TA?

- Methodology :
  - NYSE US 100 Stocks – daily data – 2000-2006
  - TA rules induced using Genetic Programming

Estimating entropy

- Active field of research in neuroscience
- Maximum-likehood ("Plug-in"):
  - empirical distribution of fixed length word
  - not suited to capture long/medium term dependencies
- Compression-based techniques :
  - Lempel-Ziv algorithm, Context-Tree Weighting
  - fast convergence rate – suited to long/medium term dependencies

Selected estimator

- Kontoyannis et al 1998

$$\hat{h}_{SM} = \left(\frac{1}{n} \sum_{i=1}^{n} \Lambda_i \right)^{-1} \log_2 n$$

Λ_i : length of the shortest string that does not appear in the *i* previous symbols

$$\downarrow$$

Example: 0 1 1 0 0 1 0 1 1 0 0

$$\Lambda_6 = 3$$

Performance of the estimator

- Experiments :
  - Uniformly distributed r.v. in {1,2,..,8} – theoretical entropy $= -\sum_{i=1}^{8} 1/8 \log_2 p = 3 \; b.p.c.$
  - Boost C++ random generator
  - Sample of size 10000

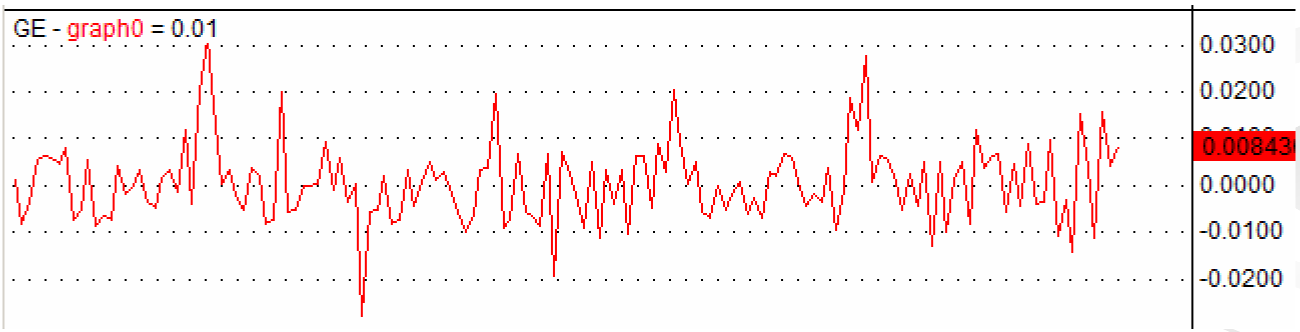
$$\hat{h}_{SM} = 2.96$$

➢ Note 1 : with sample of size 100000, $\hat{h}_{SM} \geq 2.99$

➢ Note 2 : with standard C *rand()* function and sample size = 10000, $\hat{h}_{SM} = 2.77$

Preprocessing the data (1/2)

- Log ratio between closing prices: $r_t = ln(\frac{p_t}{p_{t-1}})$
- Discretization : $\{r_t\} \in \mathbb{R} \rightarrow \{A_t\} \in \mathbb{N}$



3,4,1,0,2,6,2,...

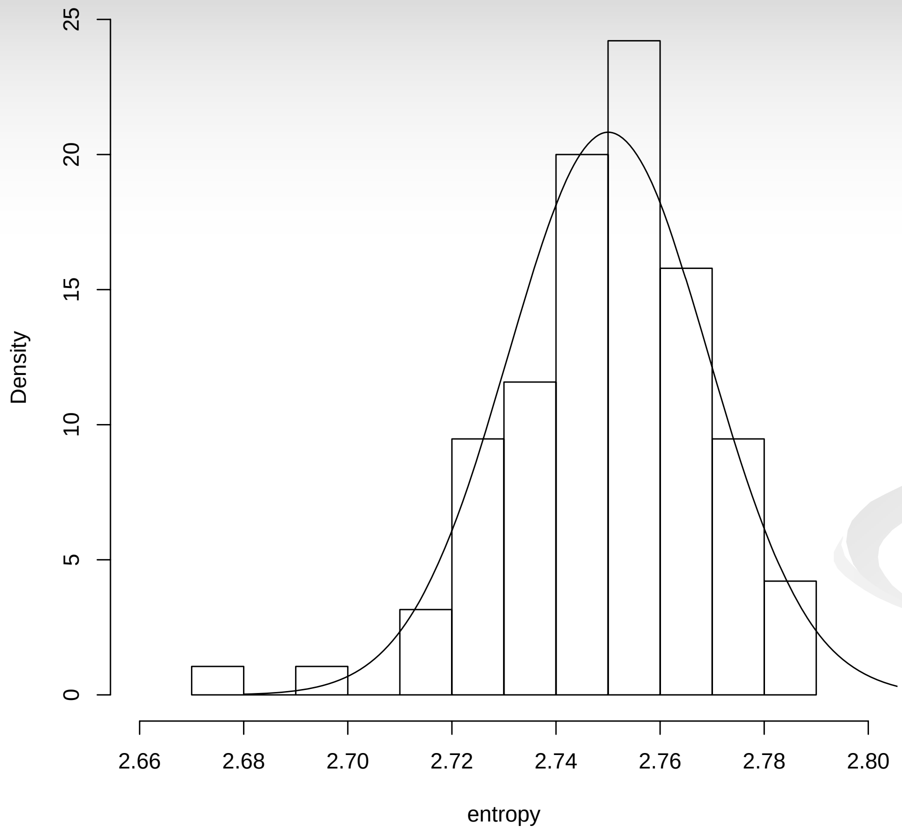
Preprocessing the data (2/2)

- Discretization is tricky – 2 problems:
  - How many bins? (size of the alphabet)
  - How many values in each bin?

- Guideline : maximize entropy with a number of bins in link with the sample size

- Here :
  - alphabet of size 8
  - same number of values in each bin ("homogeneous partitioning")

Entropy of NYSE US 100 stocks – period 2000-2006



Mean = Median = 2.75

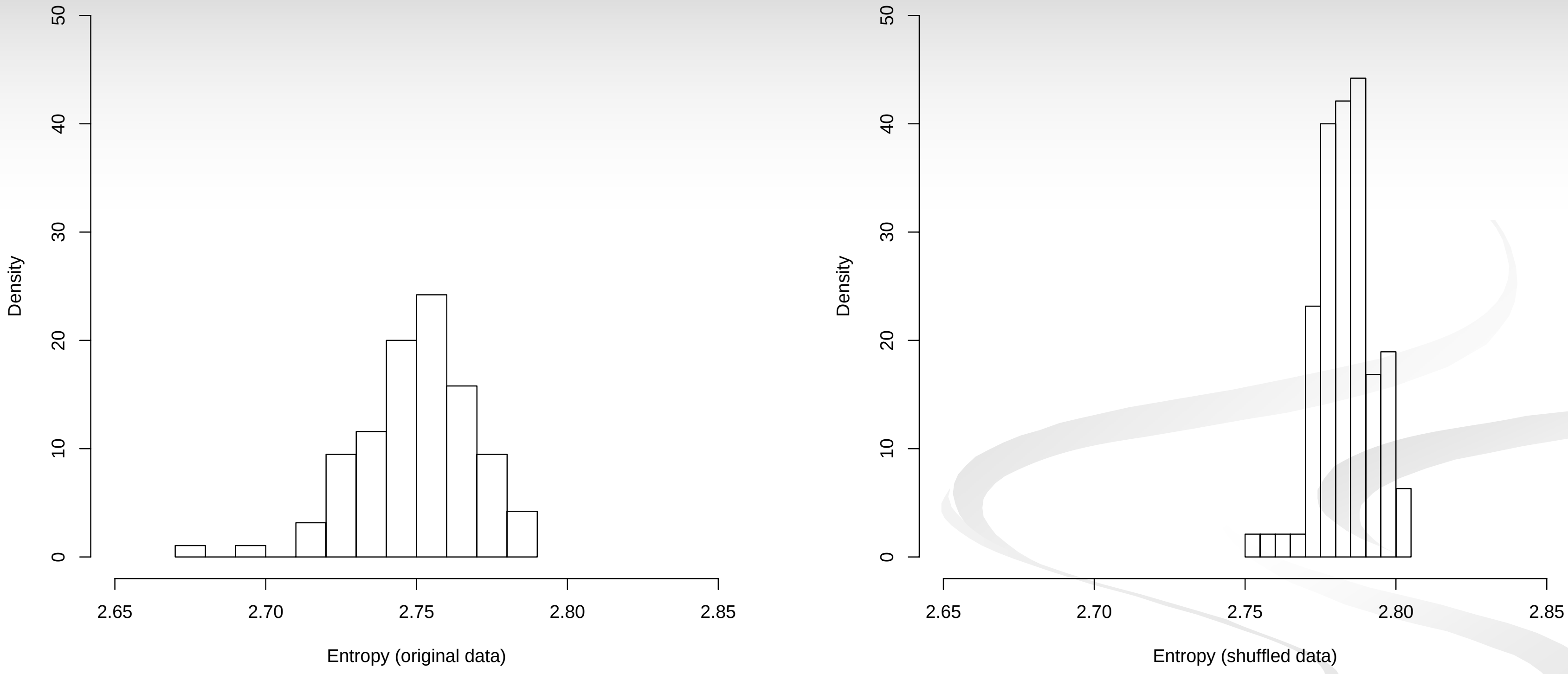
Max = 2.79

Min = 2.68

Rand() boost = 2.9

NB : a normal distribution of same mean and standard deviation is plotted for comparison.

Entropy is high but price time series are not random!



Original time series

Randomly shuffled time series

Stocks under study

Highest entropy time series

Symbol	Entropy
OXY	2.789
VLO	2.787
MRO	2.785
BAX	2.78
WAG	2.776

Lowest entropy time series

Symbol	Entropy
TWX	2.677
EMC	2.694
C	2.712
JPM	2.716
GE	2.723

BDS tests: are daily log price changes i.i.d ?

Lowest entropy time series

m	δ	TWX	EMC	C	JPM	GE
2	1	18.06	14.21	13.9	11.82	11.67
3	1	22.67	19.54	18.76	16.46	16.34
5	1	34.18	29.17	28.12	26.80	24.21

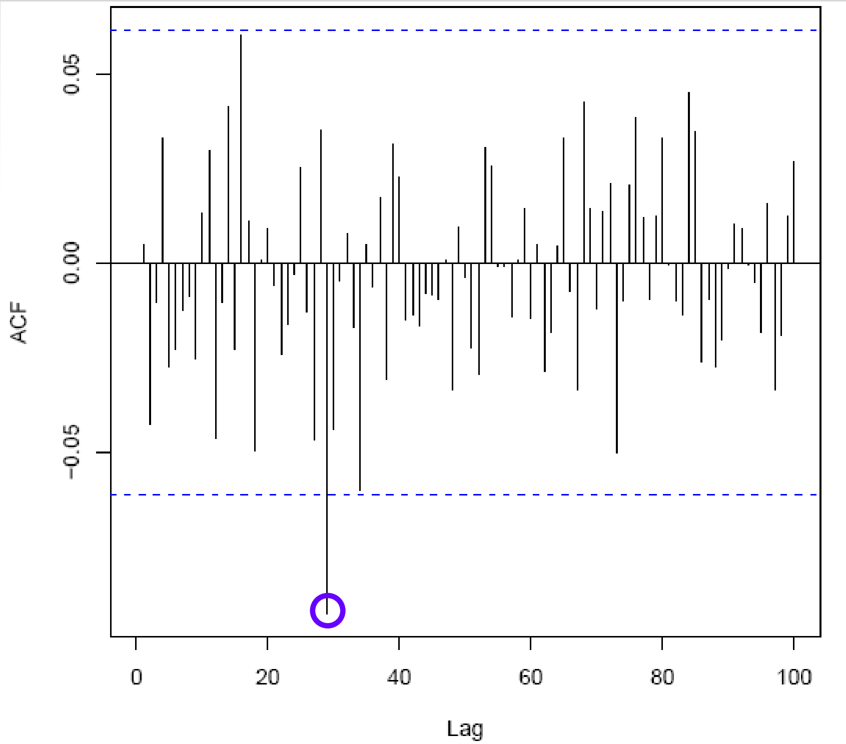
Highest entropy time series

m	δ	OXY	VLO	MRO	BAX	WAG
2	1	5.66	4.17	6.69	8.13	7.45
3	1	6.61	5.35	9.40	11.11	8.89
5	1	9.04	6.88	13.08	15.31	11.17

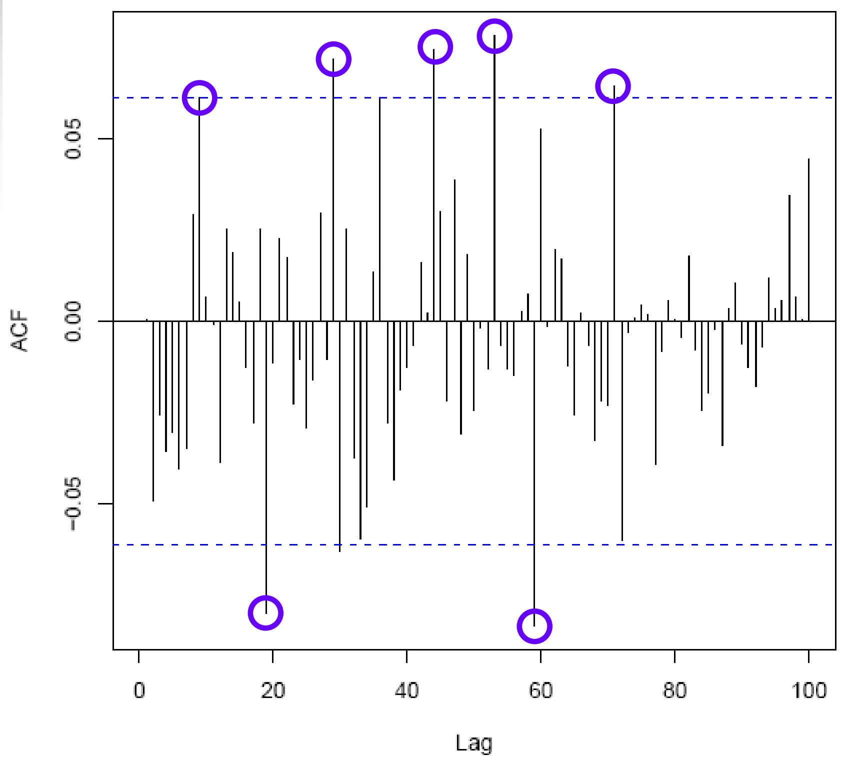
- Null that log price changes are i.i.d. always rejected at 1% level but - whatever BDS parameters - rejection is much stronger for high-entropy stocks

Autocorrelation analysis

High entropy stock (OXY)



Low entropy stock (C)

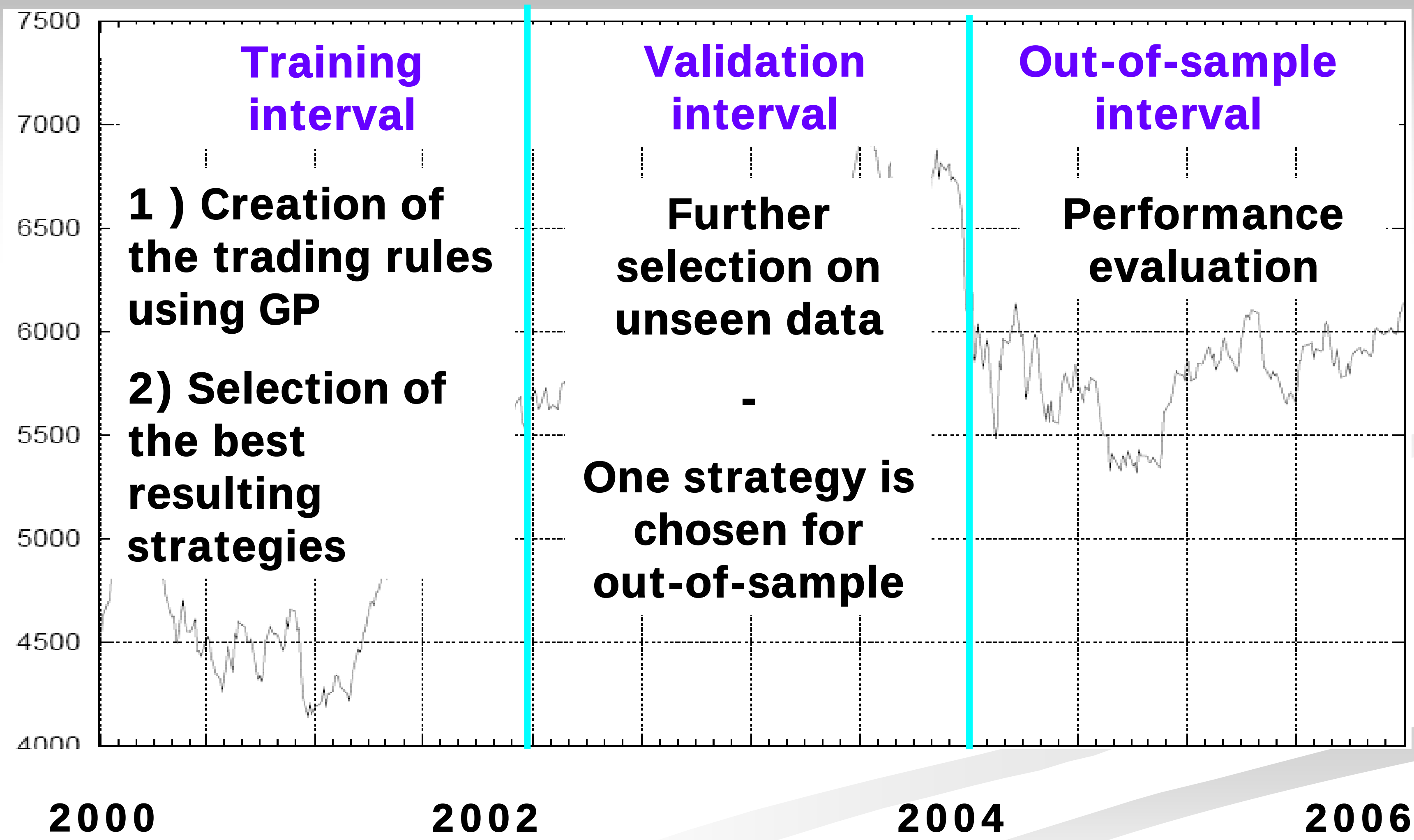


- Up to a lag 100, there are 2.7 x more autocorrelations outside the 99% confidence bands for the lowest entropy stocks than for the highest entropy stocks

Part 2 : does low entropy imply better profitability of TA?

Addressed here: are GP-induced rules more efficient on low-entropy stocks ?

GP : the big picture



GP performance assessment

- Buy and Hold is not a good benchmark
- GP is compared with lottery trading (LT) of
  - same frequency : avg nb of transactions
  - same intensity : time during which a position is held
- Implementation of LT: random sequences with the right characteristics, e.g: 0,0,1,1,1,0,0,0,0,0,1,1,1,1,1,1,0,0,1,1,0,1,0,0,0,0,0,0,1,1,1,1,1,1,…
- GP>LT ? LT>GP ? Student's t-test at 95% confidence level – 20 GP runs / 1000 LT runs

Experimental setup

- Data preprocessed with 100-days MA
- Trading systems:
  - Entry (long): GP induced rule with a classical set of functions / terminals
  - Exit:
    - Stop loss : 5%
    - Profit target : 10%
    - 90-days stop
- Fitness: net return - Initial equity: 100K$ - position sizing : 100%

Results: high entropy stocks

	GP net profits	LT net profits	GP>LT?	LT>GP?
OXY	$15.5K\$$	$14K\$$	No	No
VLO	$7K\$$	$11.5K\$$	No	No
MRO	$15K\$$	$18.5K\$$	No	No
BAX	$24K\$$	$13K\$$	**Yes**	No
WAG	$6K\$$	$-0.5K\$$	**Yes**	No

GP is always profitable

LT is never better than GP (at a 95% confidence level)

GP outperforms LT 2 times out of 5 (at a 95% confidence level)

Results: low entropy stocks

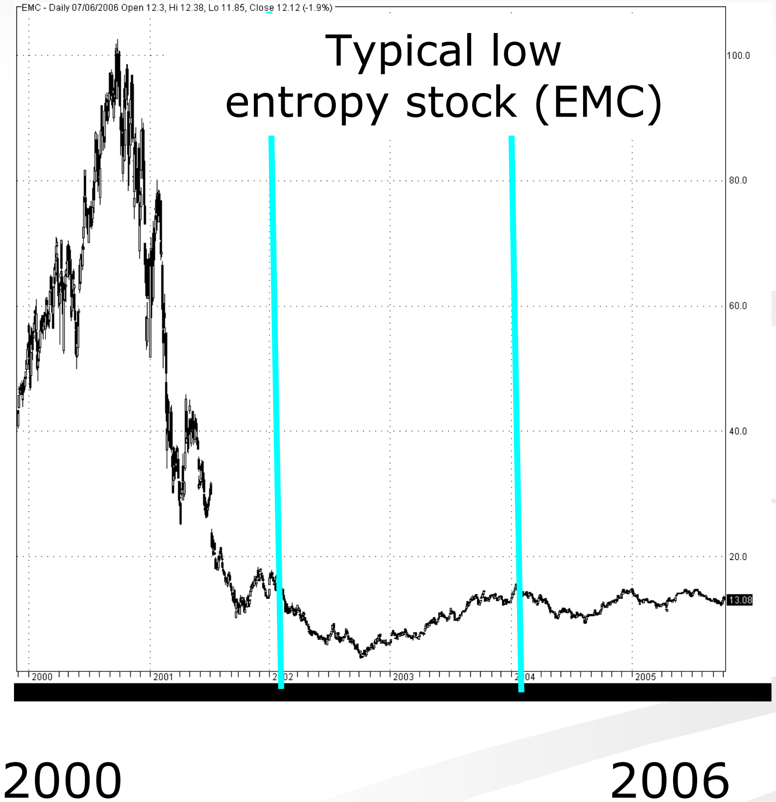
	GP net profits	LT net profits	GP>LT?	LT>GP?
TWX	$-9K\$$	$-1.5K\$$	No	**Yes**
EMC	$-16.5K\$$	$-11K\$$	No	**Yes**
C	$15K\$$	$18.5K\$$	No	No
JPM	$6K\$$	$10K\$$	No	No
GE	$-0.5K\$$	$0.5K\$$	No	No

GP is never better than LT (at a 95% confidence level)

LT outperforms GP 2 times out of 5 (at a 95% confidence level)

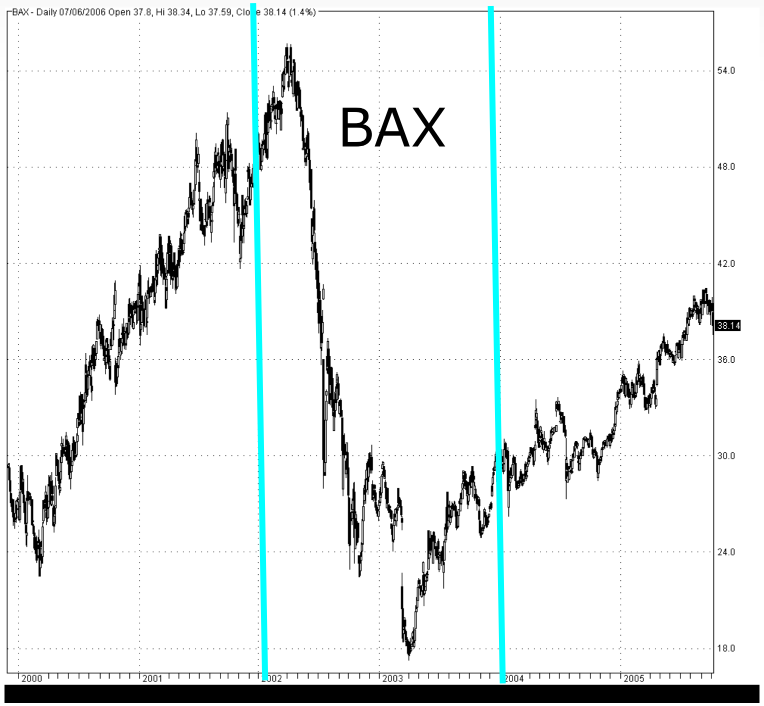
Explanations (1/2)

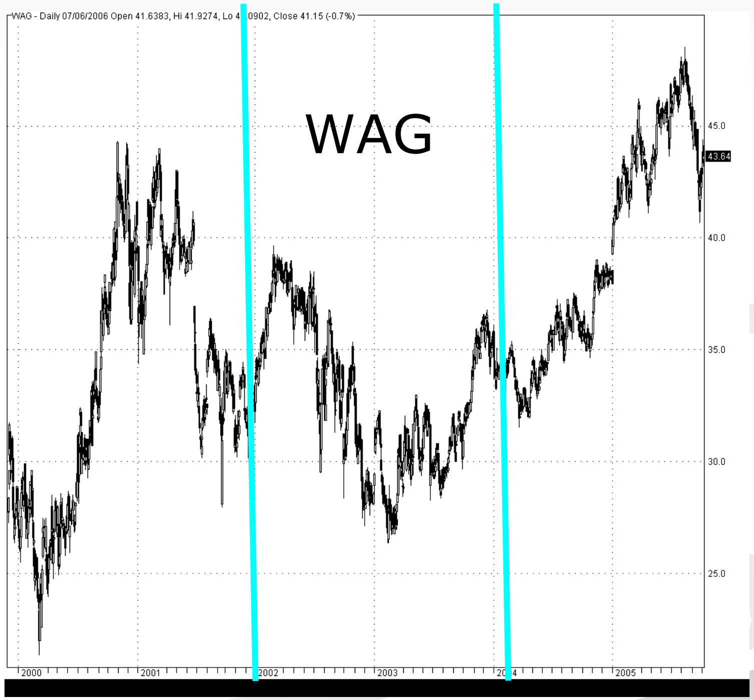
- GP is not good when training period is very different from out-of-sample e.g.



Explanations (2 / 2)

- The 2 cases where GP outperforms LT : training quite similar to out-of-sample



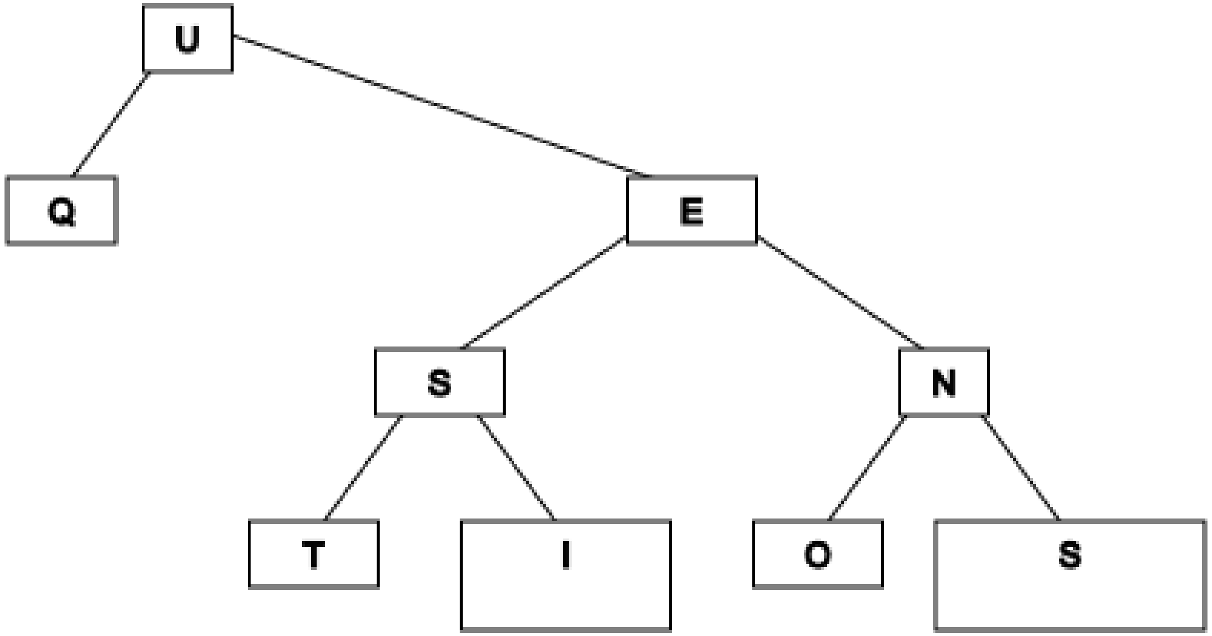


Conclusions

- EOD NYSE time series have high but differing entropies
- There are (weak) temporal dependencies
- Here, more predictable ≠ less risks
- GP works well if training is similar to out-of-sample

Perspectives

- Higher predictability level can be observed at intraday timeframe (what about higher timeframes?)
- Experiments needed with stocks less similar than the ones from NYSE US 100
- Predictability tells us about the existence of temporal patterns – but how easy / difficult to discover them ??


U
Q
E
S
N
T
I
O
S