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-likelihood (“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

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

\( \Lambda_i \): length of the shortest string that does not appear in the \( i \) previous symbols

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 \( h = -\sum_{i=1}^{8} 1/8 \log_2 p = 3 \) 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 \textit{rand()} function and sample size = 10000, \( \hat{h}_{SM} = 2.77 \)
Preprocessing the data (1/2)

- Log ratio between closing prices: \( r_t = \ln\left(\frac{p_t}{p_{t-1}}\right) \)

- Discretization: \( \{r_t\} \in \mathbb{R} \rightarrow \{A_t\} \in \mathbb{N} \)

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

GE - graph0 = 0.01
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**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>OXY</td>
<td>2.789</td>
</tr>
<tr>
<td>VLO</td>
<td>2.787</td>
</tr>
<tr>
<td>MRO</td>
<td>2.785</td>
</tr>
<tr>
<td>BAX</td>
<td>2.78</td>
</tr>
<tr>
<td>WAG</td>
<td>2.776</td>
</tr>
</tbody>
</table>

**Lowest entropy time series**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWX</td>
<td>2.677</td>
</tr>
<tr>
<td>EMC</td>
<td>2.694</td>
</tr>
<tr>
<td>C</td>
<td>2.712</td>
</tr>
<tr>
<td>JPM</td>
<td>2.716</td>
</tr>
<tr>
<td>GE</td>
<td>2.723</td>
</tr>
</tbody>
</table>
BDS tests: are daily log price changes i.i.d?

**Lowest entropy time series**

<table>
<thead>
<tr>
<th>$m$</th>
<th>$\delta$</th>
<th>$TWX$</th>
<th>$EMC$</th>
<th>$C$</th>
<th>$JPM$</th>
<th>$GE$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>18.06</td>
<td>14.21</td>
<td>13.9</td>
<td>11.82</td>
<td>11.67</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>22.67</td>
<td>19.54</td>
<td>18.76</td>
<td>16.46</td>
<td>16.34</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>34.18</td>
<td>29.17</td>
<td>28.12</td>
<td>26.80</td>
<td>24.21</td>
</tr>
</tbody>
</table>

**Highest entropy time series**

<table>
<thead>
<tr>
<th>$m$</th>
<th>$\delta$</th>
<th>$OXY$</th>
<th>$VLO$</th>
<th>$MRO$</th>
<th>$BAX$</th>
<th>$WAG$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>5.66</td>
<td>4.17</td>
<td>6.69</td>
<td>8.13</td>
<td>7.45</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>6.61</td>
<td>5.35</td>
<td>9.40</td>
<td>11.11</td>
<td>8.89</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>9.04</td>
<td>6.88</td>
<td>13.08</td>
<td>15.31</td>
<td>11.17</td>
</tr>
</tbody>
</table>

- 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

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

1) Creation of the trading rules using GP

2) Selection of the best resulting strategies

Further selection on unseen data

One strategy is chosen for out-of-sample performance evaluation

Training interval

Validation interval

Out-of-sample interval

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.: 0,0,1,1,1,0,0,0,1,1,1,1,0,0,1,1,0,1,0,0,0,0,0,1,1,1,1,1,…
- 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

<table>
<thead>
<tr>
<th></th>
<th>GP net profits</th>
<th>LT net profits</th>
<th>GP&gt;LT?</th>
<th>LT&gt;GP?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OXY</strong></td>
<td>15.5K$</td>
<td>14K$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>VLO</strong></td>
<td>7K$</td>
<td>11.5K$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>MRO</strong></td>
<td>15K$</td>
<td>18.5K$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>BAX</strong></td>
<td>24K$</td>
<td>13K$</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>WAG</strong></td>
<td>6K$</td>
<td>−0.5K$</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>GP net profits</th>
<th>LT net profits</th>
<th>GP&gt;LT?</th>
<th>LT&gt;GP?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TWX</strong></td>
<td>$-9K$</td>
<td>$-1.5K$</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>EMC</strong></td>
<td>$-16.5K$</td>
<td>$-11K$</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>$15K$</td>
<td>$18.5K$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>JPM</strong></td>
<td>$6K$</td>
<td>$10K$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>GE</strong></td>
<td>$-0.5K$</td>
<td>$0.5K$</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

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.

Typical low entropy stock (EMC)

2000 2006
Explanations (2/2)

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

![Graph of BAX and WAG stocks](image.png)
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 ??