Entropy Rate and Profitability of Technical Analysis: Experiments on the NYSE US 100 Stocks

Nicolas NAVET – INRIA
France
nnavet@loria.fr

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

CIEF2007 - 07/22/2007

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”):
  - construct an n-th order Markov chain
  - 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 previous symbols

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

\[ \Lambda_6 = 3 \]
Performance of the estimator

- **Experiments**: 
  - Uniformly distributed r.v. in \{1,2,..,8\} – theoretical entropy: \[-\sum_{i=1}^{8} \frac{1}{8} \log_2 p = 3 \text{ b.p.c.}\]
  - 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: \(r_t = \ln\left(\frac{p_t}{p_{t-1}}\right)\)
- Discretization: \(\{r_t\} \in \mathbb{R} \rightarrow \{A_t\} \in \mathbb{N}\)

3,4,1,0,2,6,2,...
Preprocessing the data (2/2)

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

Entropy of NYSE US 100 stocks – period 2000-2006

Mean = Median = 2.75
Max = 2.79
Min = 2.68
Rand() C lib = 2.77 !

Note: 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

<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>
<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>

Highest entropy time series

Lowest entropy time series
Autocorrelation analysis

- Up to a lag 100, there are on average 6 autocorrelations outside the 99% confidence bands for the lowest entropy stocks versus 2 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

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:
  0,0,1,1,1,0,0,0,0,1,1,1,1,1,0,0,1,1,1,0,0,0,0,0,0,0,0,1,1,1,1,...
- 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>OXY</td>
<td>15.5K$</td>
<td>14K$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>VLO</td>
<td>7K$</td>
<td>11.5K$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>MRO</td>
<td>15K$</td>
<td>18.5K$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>BAX</td>
<td>24K$</td>
<td>13K$</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>WAG</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>TWX</td>
<td>−9K$</td>
<td>−1.5K$</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>EMC</td>
<td>−16.5K$</td>
<td>−11K$</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>C</td>
<td>15K$</td>
<td>18.5K$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>JPM</td>
<td>6K$</td>
<td>10K$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>GE</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.
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??