Predicting a Rank Measure for Portfolio Selection

The Third International School on
Actuarial and Financial Mathematics
Feodisia Ukraine
5th of September 2000

Thomas Hellström
University of Umeå
Sweden

http://www.cs.umu.se/~thomash/
email: thomash@cs.umu.se

Overview of the talk

- An alternative formulation of the stock prediction problem
- Statistical investigation of the ranks
- Predicting ranks with linear models
- Evaluation as time series predictions and by Simulated Trading
- Excuses for the too good results…
- Conclusions

Individual Stock Returns R(t)

The usual approach is to attempt to predict the returns R(t) for a stock's Close prices:

$$R_k(t) = \frac{Close(t)\text{-}Close(t-k)}{Close(t-k)} \approx \log\frac{Close(t)}{Close(t-k)}$$

- This is a well known hard problem. to say the least…
- It is not even a complete solution since we want to SELECT among many stocks. Picking the stock with highest predicted return is not necessarily optimal.

Ranking Stock Returns R(t)

For a stock m in a set of N stocks, the h-day rank for day t is defined as

$$A_h^m(t) \equiv \frac{\#\{R_h^i(t) \mid R_h^m(t) \geq R_h^i(t), 1 \leq i \leq N\} - 1}{N-1} - 0.5$$

- The stock with highest R gets rank 0.5
- The stock with lowest R gets rank -0.5
- The median stock gets rank 0

Nice things about ranks:

- Predicting the rank is (at least) as good as predicting the returns.
- Clear benchmark: hit rate for sign > 50%.
- Uniform distribution.
- The effect of global events gets automatically incorporated.

One-Day Memory in the Ranks

$A_1(t+1)$ tabulated as a function of $A_1(t)$
207 Swedish stocks 1987-1997

A (t)

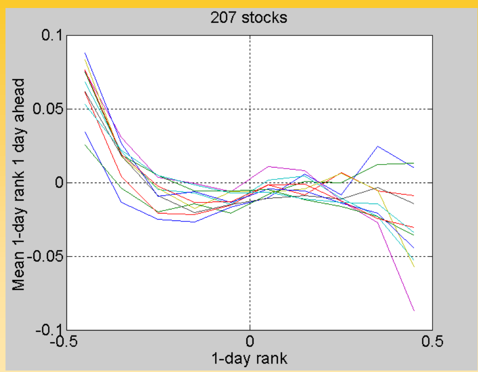
A(t+1)	-0.45	-0.35	-0.25	-0.15	-0.05	0.05	0.15	0.25	0.35	0.45
Fraction% >0	**59.4**	52.9	49.1	47.3	48	49.6	49.5	48.2	47.8	46.4
Mean rank A	0.067	0.017	-0.005	-0.011	-0.011	-0.004	-0.005	-0.01	-0.014	-0.033
no. of obs.	30878	30866	31685	30837	30434	31009	31258	30539	30951	31550

With 59.4% probability:

- The worst performing 10% of the stocks will be in the upper half next day.
- This prediction can be done EVERY day (since there is always a worst performing 10%).

One-Day Memory in the Ranks

one curve per year 1987-1997



data from the Swedish stock market

Five-Day Memory in the Ranks

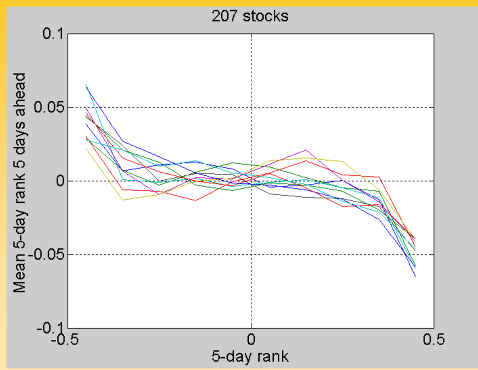
$A_5(t+5)$ tabulated as a function of $A_5(t)$
207 Swedish stocks 1987-1997

	A (t)									
A(t+5)	-0.45	-0.35	-0.25	-0.15	-0.05	0.05	0.15	0.25	0.35	0.45
Fraction% >0	56.5	51.9	50.1	50.3	50	50.2	49.4	48.4	48.4	44.3
Mean rank A	0.04	0.01	0	0	0	0	0	-0.01	-0.01	-0.04
no. of obs.	30692	30689	31485	30652	30286	30772	30955	30349	30644	31026

With 56.5% probability:
- ✔ The worst performing 10% of the stocks will be in the upper half next day.
- ✔ This prediction can be done EVERY day (since there is always a worst performing 10%).

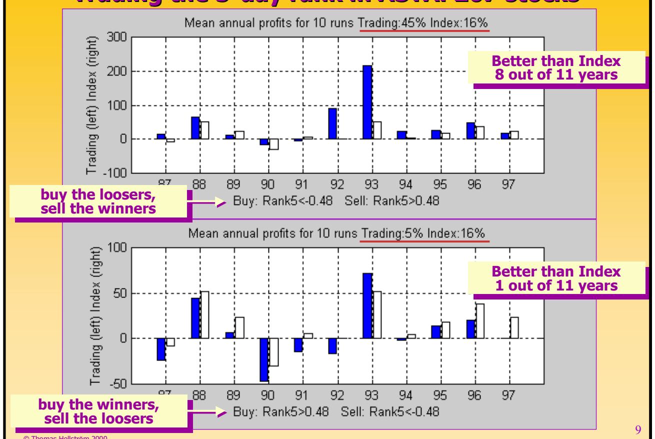
Five-Day Memory in the Ranks

one curve per year 1987-1997

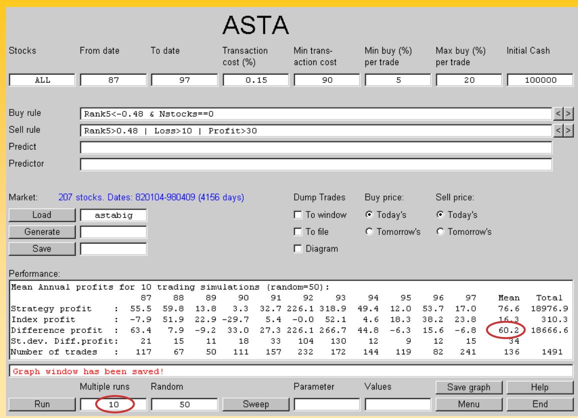


data from the Swedish stock market

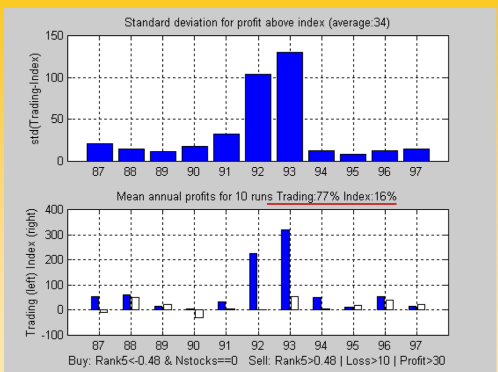
Trading the 5-day rank in ASTA. 207 stocks



Refined trading rules with 5-day rank.



Refined trading rules with 5-day rank.



How About Predicting Ranks?

Use the rank variables as inputs and/or outputs in a time series prediction.

Notation:

The h-day rank for stock m on day t : $A_h^m(t)$

The h-day prediction (computed at time t) : $\hat{A}_h^m(t+h)$

Model:

$$\hat{A}_h^m(t+h) = g_m(A_1^m(t), A_2^m(t), A_5^m(t), A_{20}^m(t))$$

A model for Prediction of the Ranks

Let's try a linear model:

$$\hat{A}_h^m(t+h) = p_0 + p_1 A_1^m(t) + p_2 A_2^m(t) + p_3 A_5^m(t) + p_4 A_{20}^m(t)$$

I.e.: The h-day rank at time t+h is predicted from the 1-day, 2-day, 5-day and 20-day ranks computed at time t.

For a market with N stocks we build N models ($1 \quad m \quad N$).
To facilitate comparison the m predictions are scaled similar to the rank definition:

$$\hat{A}_h^m(t+h) \leftarrow \frac{\#\{\hat{A}_h^i(t) \mid \hat{A}_h^m(t) \geq \hat{A}_h^i(t), 1 \leq i \leq N\} - 1}{N-1} - 0.5$$

The parameter vector $(p_0, p_1, p_2, p_3, p_4)$ can be determined by linear regression on historical data.

Experimental Set-Up

- One-day predictions of one-day ranks $\hat{A}_1^m(t+1)$
- 80 Swedish stocks from 1989-1997
- Sliding windows:
  1000 days for modelling and the following 100 for prediction. Step 100 days and repeat until end of data.

1000 days for modelling (computation of $(p_0, p_1, p_2, p_3, p_4)$) | 100 days for predictions | *time*

- I.e.: $\hat{A}_1^m(t+1)$ is based on close prices up to time t:

$$\hat{A}_1^m(t+1) = p_0 + p_1 A_1^m(t) + p_2 A_2^m(t) + p_3 A_5^m(t) + p_4 A_{20}^m(t)$$

Results

52.8% of all predictions >0 result in a rank >0

Table 4: 1-day predictions of 1-day ranks $|\hat{A}_1(t+1)| > 0.00$.

$Year:$	93	94	95	96	97	93-97
$Hitrate_+$	51.3	53.4	53.1	53.1	53.0	52.8
$Hitrate_-$	51.9	53.6	53.1	53.4	53.1	53.0
$Meanrank_+$	0.011	0.025	0.022	0.021	0.018	0.020
$Meanrank_-$	-0.014	-0.025	-0.021	-0.023	-0.019	-0.020
$Return_+$	0.396	0.101	0.139	0.253	0.190	0.217
$Return_-$	0.247	-0.171	-0.081	0.042	-0.009	0.006
$Return_{tot}$	0.391	0.027	0.034	0.143	0.111	0.142
$\#Pred_+$	7715	8321	8311	8923	8162	41506
$\#Pred_-$	7788	8343	8342	8942	8171	41664
$\#Pred$	15503	16664	16653	17865	16333	83170

The mean return after a prediction >0 is 0.217%
The mean return after a prediction <0 is 0.006%

Selected Predictions

63% of predictions >0.49 result in a rank >0

Table 6: 1-day predictions of 1-day ranks $|\hat{A}_1(t+1)| > 0.49$.

$Year:$	93	94	95	96	97	93-97
$Hitrate_+$	57.7	67.3	63.3	65.7	60.6	63.0
$Hitrate_-$	51.8	57.0	56.4	57.5	55.5	55.7
$Meanrank_+$	0.061	0.115	0.088	0.115	0.077	0.092
$Meanrank_-$	-0.007	-0.070	-0.069	-0.046	-0.034	-0.045
$Return_+$	1.202	0.841	0.618	0.726	0.686	0.827
$Return_-$	0.801	-0.490	-0.332	0.073	-0.064	0.003
$Return_{tot}$	0.391	0.027	0.034	0.143	0.111	0.142
$\#Pred_+$	215	214	218	230	213	1092
$\#Pred_-$	220	221	220	233	218	1114
$\#Pred$	15503	16664	16653	17865	16333	83170

The mean return after a prediction >0.49 is 0.827%
The mean return after a prediction <-0.49 is 0.003%

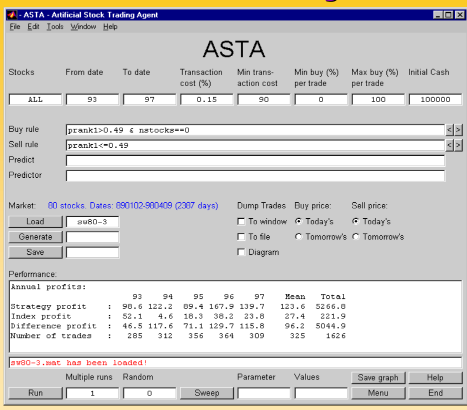
Simulated Trading

We are using ASTA to execute the following trading strategy:

Buy rule:	prank1>0.49 & nstocks==0
Sell rule:	prank1<=0.49

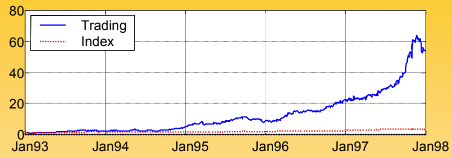
- The function prank1 implements the described one-day predictions $\hat{A}_1^m(t+1)$.
- We are buying the predicted 1% best performing stock(s) every day if they are not already in the portfolio.
- We are selling every day if the stock wont generate a buy signal again.
- Transaction costs: 0.15%

Simulated Trading

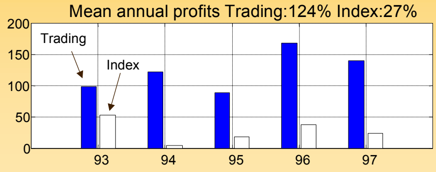


Simulated Trading

Equity curves for Trading (5267%) and Index (222%)



Mean annual profits Trading:124% Index:27%



The Trading is consistently better than Index every year !

Buy: prank1>0.49 & nstocks==0   Sell: prank1<=0.49

Statistics for the Simulated Trading

- The number of trades is high. This increases the statistical significance.
- "Only" 56.2% profitable trades seems to be enough to generate a huge profit

Number of sell orders	812
Mean profit per trade	1.2%
Trades with profit>0	456 =56.2%
Trades with profit<0	212 =26.1%
Trades with profit=0	144=17.7%
Total Increase	5266.8 %
Total increase for index	221.9 %
Annual Profit	123.6%
Annual Profit for index	27.4%
Median Excess profit	115.8%
Number of ignored buy signals	148 (15.4%)

Where Did We Cheat ?

The results actually look too good…

Is this a clear example of market inefficiency and a refutation of the Efficient Market Hypothesis ?

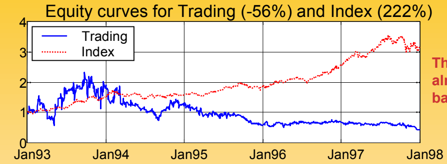
Possible "explanation":

- The prediction $\hat{A}_1^m(t+1)$ is based on close prices y: $y_m(t-k), \ldots, y_m(t)$. I.e.: The trades performed on day t assume knowledge of the close prices for day t.
  - This is not possible!
  - However, it is very often ignored when one-day predictions are evaluated.
- Another explanation: The excess profit for the trading strategy is paid by increased risk (this argument can always be used)

Simulated Trading

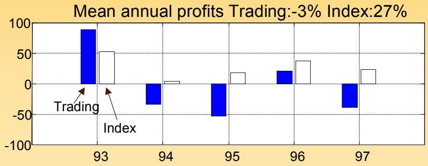
The predictions use stock prices up to the day BEFORE the portfolio rebalancing. (this is not very realistic either….)

Equity curves for Trading (-56%) and Index (222%)



The Trader almost goes bankrupt !

Mean annual profits Trading:-3% Index:27%



Buy: prank1>0.49 & nstocks==0   Sell: prank1<=0.49

Conclusions

- A real evaluation of the trading strategy has to involve open prices or intra-day data to be realistic.
- It indicates market inefficiencies and casts doubts on the Efficient Market Hypothesis.
- The shown rank predictions are in sharp contradiction with the Random Walk Hypothesis for stock prices. We are able to predict the sign of the rank consistently over a 5-year-period of daily predictions.
- The general idea of predicting ranks instead of individual returns seems to be successful.
- Non linear rank models such as Neural Networks is an interesting topic for future research.

- References:
  - Predicting a Rank measure for Portfolio Selection

    Thomas Hellström. Technical report UMINF 00.18
    Department of Computing Science, Umeå University
  - ASTA - User's Reference Guide

    Thomas Hellström. Technical report UMINF 00.16
    Department of Computing Science, Umeå University

- ASTA is written in MATLAB and is available for download for academic use. More information on http://www.cs.umu.se/~thomash/asta/home