FE5209 Financial Econometrics Group Project Report

Tutorial Group Alpha

HO Ngok Chao(Algorithm, 20%), HU Mingyue(Algorithm, 20%)

ZHU Haolin(Data, 20%), Feng Tiantian(Analysis, 20%), SU Yulu(Analysis, 20%)

Introduction

Our project is about active investment strategies. The existing literature analyzed the performance of various strategies in ancient and small datasets. Therefore, our group tested the performance of those strategies in newer and larger datasets and compare their performance with and without transaction cost.

All strategies we implemented are packaged as function and made available at https://gitlab.com/johnho/rmr.

Methodology of RMR and ARIMA

Portfolio selection strategies we analyzed are all self-financed with no margin/short-sale allowed. Then, portfolio selection is about assigning positive weight to each asset.

Assuming there are d assets available in the market, on the t_{th} period, **close price vector** is obtained as $P_t \in R_+^d$ .Then, the **relative price vector** $x_t = (x_t^1, \dots, x_t^d) \in R_+^d$ is computed as $x_t^i = p_t^i / p_{t-1}^i$. Next, the **portfolio vector** is designed according to the strategy: $b_t = (b_t^1, \dots, b_t^d)$. At last, given initial wealth is $\boldsymbol{S_0}$, after n trading period, the portfolio cumulative wealth is $\boldsymbol{S_n}(\boldsymbol{b_1^n}, \boldsymbol{x_1^n}) = \boldsymbol{S_0} \prod_{\boldsymbol{t}}^{\boldsymbol{T}} (\boldsymbol{b_t^T} \boldsymbol{x_t})$.

The ancestor of RMR is basic mean reversion strategy. The problems of mean reversion are single-period mean reversion assumption is not always satisfied in the real world and when Data contain a lot of noise and outliers and thus substantially influences the effectiveness of the algorithm (Huang, 2016).

To address these problems, RMR is derived from **two optimization problems.** The 1st optimization problem is about predicting next price. Robust L_1-Median Estimator at the end of t^{th} period is $\tilde{P}_{t+1} = L_1 med_{t+1}(\omega) = \mu$, where ω is the window size, μ denotes the L_1-Median Estimator optimal value of the optimization problem $\mu = argmin_\mu \sum_{i=0}^{k-1} \|P_{t-i} - \mu\|$, where $\| . \|$denotes the Euclidean norm and then $\tilde{x}_{t+1}(\omega) = \frac{L_1\, med_{t+1}(\omega)}{P_t} = \frac{\mu}{P_t}$.

The 2nd optimization problem is to find an optimal portfolio by minimizing the deviation from last portfolio b_t under the condition of $b^T \tilde{x}_{t+1} \geq \varepsilon$, expressed as $b_{t+1} = \ argmin_{b \in \Delta d} \frac{1}{2} \|b - b_t\|^2 \ s.t. \ \ b^T \tilde{x}_{t+1} \geq \varepsilon.$

The 1st problem can be solved iteratively using function T (Left) and the 2nd problem can be solved using Lagrange Multiplier (Right) (Huang, 2016).

Proposition 1. *The solution of L_1-median-MAADM optimization problem 1 is calculated through iteration, and the iteration process is described as:*

$$\boldsymbol{\mu} \rightarrow T(\boldsymbol{\mu}) = \left(1 - \frac{\eta(\boldsymbol{\mu})}{\gamma(\boldsymbol{\mu})}\right)^{+} \tilde{T}(\boldsymbol{\mu}) + \min\left(1, \frac{\eta(\boldsymbol{\mu})}{\gamma(\boldsymbol{\mu})}\right) \boldsymbol{\mu},$$

where

$$\eta(\boldsymbol{\mu}) = \begin{cases} 1 & if \ \boldsymbol{\mu} = \mathbf{p}_{t-i}, \quad i = 0, \ldots, k-1 \\ 0 & otherwise \end{cases},$$
$$\gamma(\boldsymbol{\mu}) = \|\tilde{R}(\boldsymbol{\mu})\|, \quad \tilde{R}(\boldsymbol{\mu}) = \textstyle\sum_{\mathbf{p}_{t-i} \neq \boldsymbol{\mu}} \frac{\mathbf{p}_{t-i} - \boldsymbol{\mu}}{\|\mathbf{p}_{t-i} - \boldsymbol{\mu}\|},$$
$$\tilde{T}(\boldsymbol{\mu}) = \left\{\textstyle\sum_{\mathbf{p}_{t-i} \neq \boldsymbol{\mu}} \frac{1}{\|\mathbf{p}_{t-i} - \boldsymbol{\mu}\|}\right\}^{-1} \textstyle\sum_{\mathbf{p}_{t-i} \neq \boldsymbol{\mu}} \frac{\mathbf{p}_{t-i}}{\|\mathbf{p}_{t-i} - \boldsymbol{\mu}\|}.$$

$$\mathbf{b}_{t+1} = \mathbf{b}_t - \alpha_{t+1}(\hat{\mathbf{x}}_{t+1} - \overline{x}_{t+1} \cdot \mathbf{1}),$$

where $\overline{x}_{t+1} = \frac{1}{d}(\mathbf{1} \cdot \hat{\mathbf{x}}_{t+1})$ denotes the average predicted price relative and α_{t+1} is the Lagrangian multiplier calculated as,

$$\alpha_{t+1} = \min\left\{0, \frac{\hat{\mathbf{x}}_{t+1}^T \mathbf{b}_t - \epsilon}{\|\hat{\mathbf{x}}_{t+1} - \overline{x}_{t+1} \cdot \mathbf{1}\|^2}\right\}.$$

As the 2nd optimization of RMR is simply a regularization for large change which will incur large transaction cost and 1st optimization is about prediction price, it is obvious we can replace Robust L_1-Median Estimator with **ARIMA** model and form a new strategy to be compared with RMR later using auto.arima().

Experiment 1

Experiment 1 is the realization of our reference paper and ARIMA. To estimate the strength of RMR strategy, four other strategies are compared in the tests.

1) **Best-stock ('BEST')**: Buys the best stock over all period including the future. It is a hindsight strategy used as an optimal reference for other strategies.

2) **Passive aggressive mean reversion ('PAMR')**: Estimates the next price relative as the inverse of last price relative. It adopts the single-period mean reversion assumption, which is not satisfied with reality.

3) **Online Moving Average Reversion ('OLMAR'):** Predicts the next price relative using moving averages and explores the multi-period mean reversion.

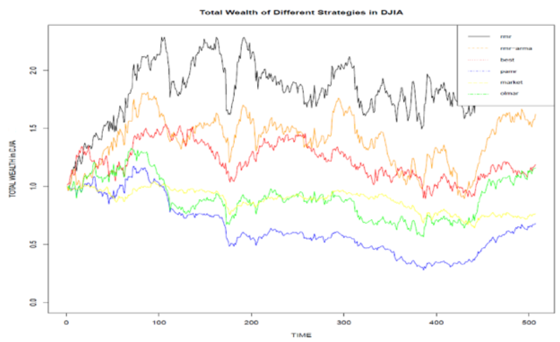
4) **Market:** Buys assets according to a pre-defined weight and holds until the end.

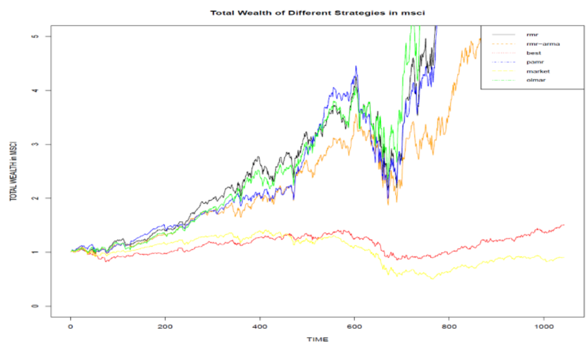
There are three assumptions which are the basic principles for the portfolio selection strategy. There is no transaction cost or taxes in this PS model; One can buy and sell required quantities at last closing price of any given trading period; Market behaviour is not affected by a PS strategy.

Identically parameters are used for each dataset. For all datasets, the length of window and sensitivity are both 5. The author only focused on the North America market and examined each market in respective sub-period.

Data set	Region	Time Frame	#days	#assets
DJIA	US	14/01/2001 - 14/01-2003	507	30
SP500	US	02/01/1998 - 31/01/2003	1276	25
TSE	CA	04/01/1994 - 31/12/1998	1259	88
MSCI	Global	01/04/2006 – 31/03/2010	1043	24
NYSE(O)	US	03/04-1962 – 31/12/1984	5651	36
NYSE(N)	US	01/01/1985 – 30/06/2010	6431	23

In experiment 1, we exploit the six datasets to draw the plots and below are two of them. First plot is the portfolio cumulative wealth of DJIA and second is from MSCI. Obviously, the RMR strategy which is the **black line** in the plot hovers over other strategies and have good results. The result is identical to our reference paper.





RMR strategy is promising and reliable PS technique to achieve high return. Compared with the existing mean reversion strategies (PAMR and OLMAR), RMR strategies obtained higher cumulative wealth in the datasets NYSE(O), NYSE(N) and DJA. The above two plots with **orange line** indicate that **ARIMA** works well.

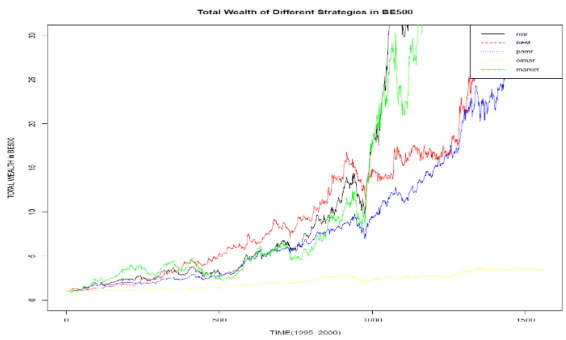
Experiment 2

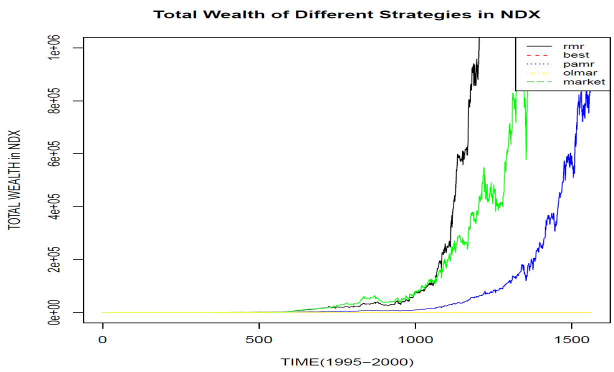
The experiment 2 is to calculate performance of these strategies based on the data we collected and innovations. Compared to the dataset used in our reference paper, our selection criterion of data is more meaningful.

Data set	Region	Time Frame	#days	#assets
BE500	UK	09/01/1995 - 01/01-2018	5993	435
IBOV	BR	04/01/1995 - 28/12/2017	5988	56

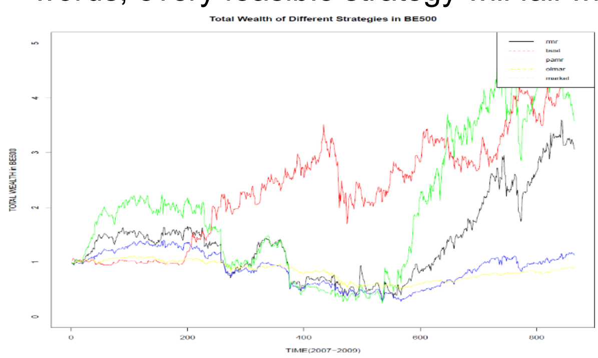
NDX	US	04/01/1995 – 29/12/2017	5995	85
NKY	JP	04/01/1995 – 29/12/2017	5990	213
TWSE	TW	06/01/1995 – 29/12/2017	5992	733

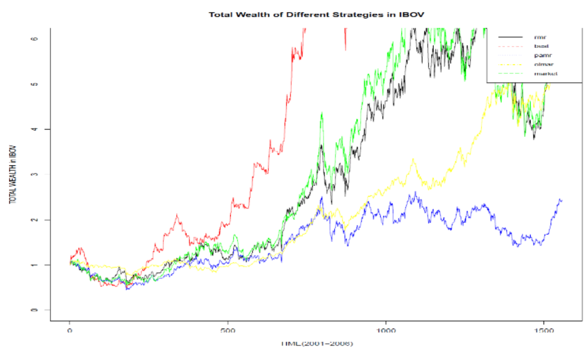
In experiment 2, we select 5 different market indices from all over the world, including Japan, Taiwan, UK and Brazil. These indices are influential in its respective continents. In our experiment 2, we divided the whole period into four, covering 1995 dot com bubble and 2007-2008 financial crisis. The aim of this kind of period segmentation is to test the performances of different portfolio strategies in both downtrend and steady economic environment.



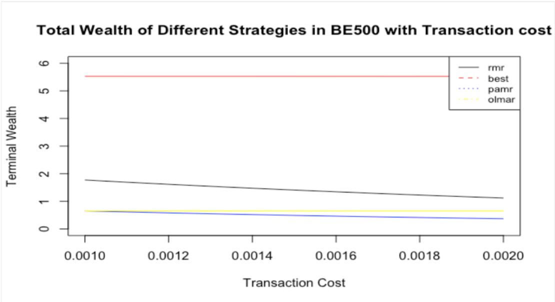


We find that RMR did well in the total wealth of IBOV-p1, BE500-p1 and NDX-p1 under steady economic environment. However, there are some exceptions. The performances of mean-reversion strategies are not so good in financial crisis in the markets except North America one. In the graph of BE500 Period 3, we can see that they are beaten by best strategy which are hindsight using future data. In other words, every feasible strategy will fail when crisis comes.





Another exception is the scenario of emerging market which doesn’t have a complete security market system. From the picture of strategy performances in Brazilian market (IBOV), mean-reversion strategies were once again beaten by best strategy. The terminal wealth is low in around 2000 and climbs slowly in next 4 years. However, after 2005 Brazilian security market regulation reform, people’s behaviour regarding Brazilian stock market changed and the performances of mean-reversion strategies started to get better. To sum up, immature security market does affect the performances of mean-reversion strategies.



BE500	Sharpe ratio
BEST	1.58071075
OLMAR	0.42439661
PAMR	-0.0574036
RMR	0.36451735

Another improvement is adding transaction cost when analyzing performance of each strategy. We realize that TC has an influence on the RMR. Following is without TC. Except for the visible plot, we also consider APY (annualized percentage yield), WT (winning ratio), Sharpe ratio, MDD (maximum drawdown) and CR (Calmar ratio) in the experiment, the results are identical to the plots. In addition, we conduct a statistical t-test to evaluate whether alpha is significantly different from 0.

		Experiment 1					Experiment 2-1					Experiment 2-2			
Criteria	Dataset	BEST	OLMAR	PAMR	RMR	Dataset	BEST	OLMAR	PAMR	RMR	Dataset	BEST	OLMAR	PAMR	RMR
APY		0.0901	0.0776	-0.1753	0.6334		1.3091	12.0080	8.7290	13.9750		0.7329	0.5293	0.0404	0.4526
WT		0.5049	0.5207	0.5187	0.5503		0.5013	0.5730	0.5973	0.5775		0.5156	0.5075	0.5168	0.5098
Sharpe Ratio	DJIA	0.1174	0.0728	-0.4637	1.1597	NDX	1.6342	12.7110	12.1172	15.3026	BE500(2007-2009)	1.5068	0.4879	0.0007	0.4589
MDD		0.4187	0.5762	0.7646	0.3469		0.6651	0.4944	0.3706	0.4133		0.5174	0.8853	0.8016	0.8322
CR		0.2152	0.1346	-0.2293	1.8258		1.9682	24.2886	23.5520	33.8171		1.4167	0.5978	0.0504	0.5439
APY		0.1074	1.1783	0.9756	1.0234		1.1019	2.6569	1.6890	2.0003		0.7882	0.4424	0.1764	0.3963
WT		0.5197	0.6002	0.5916	0.5916		0.5677	0.5469	0.6094	0.5729		0.4987	0.5160	0.4859	0.5064
Sharpe Ratio	MSCI	0.3391	2.8418	2.5647	2.5128	BE500(1995-2000)	2.3713	3.9008	5.2323	3.5157	IBOV	1.7534	0.8126	0.3260	0.7485
MDD		0.3935	0.4203	0.5528	0.4933		0.1994	0.3302	0.1033	0.2134		0.6475	0.5152	0.5903	0.4528
CR		0.2730	2.8035	1.7648	2.0747		5.5265	8.0456	16.3451	9.3727		1.2173	0.8587	0.2988	0.8753

t test	Dataset	BEST	OLMAR	PAMR	RMR	Dataset	BEST	OLMAR	PAMR	RMR	Dataset	BEST	OLMAR	PAMR	RMR
p-value	DJIA	0.0752	0.1312	0.4132	0.0049	NDX	0.0216	0.0000	0.0000	0.0000	BE500(2007-2009)	0.0060	0.0456	0.1752	0.0473
alpha		0.0012	0.0013	0.0002	0.0030		0.0020	0.0098	0.0080	0.0102		0.0023	0.0035	0.0008	0.0030
p-value	MSCI	0.1211	0.0000	0.0000	0.0000	BE500(1995-2000)	0.0296	0.0110	0.0001	0.0111	IBOV	0.0102	0.3140	0.8207	0.3864
alpha		0.0004	0.0033	0.0029	0.0030		0.0038	0.0065	0.0045	0.0052		0.0014	0.0003	-0.0005	0.0002

Conclusion

There are two main conclusions. First, RMR's performance is highly affected by transaction cost. Second, RMR performs well in most markets but there're some exceptions. During financial crisis, RMR had lower Sharpe Ratio and Calmar Ratio. Mean reversion strategy did not perform best in emerging markets which have immature market system.

Reference

D. Huang, J. Zhou, B. Li, S. Hoi, and S. Zhou, “Robust Median Reversion Strategy for Online Portfolio Selection,” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 28, NO. 9, SEPTEMBER 2016.

B. Li and S. C. H. Ho, “On-line portfolio selection with moving average reversion,” in Proc. Int. Conf. Mach. Learning, 2012, pp. 273–228.