

An On-line Machine Learning Return Prediction

Yueliang (Jacques) Lu Weidong Tian

Belk College of Business, University of North Carolina at Charlotte

October, 2020

IRMC 2020 - Global Virtual Conference

A preliminary version appears as “*Predict Relative Stock Index Returns by Sufficient Historical Data*”



UNIVERSITÀ
DEGLI STUDI
FIRENZE

NYU STERN

Salomon Center
for the Study of
Financial Institutions

FDC FUNDAÇÃO
DOM CABRAL



European
Commission

Sponsors

S&P Global
Market Intelligence



1. Propose a new prediction methodology for *relative stock index returns* based on universal portfolio construction
 - *relative stock index return* \equiv gross return of a stock index / (1 + interest rate)
 - An on-line strategy competitive with the best *constant proportional portfolio*

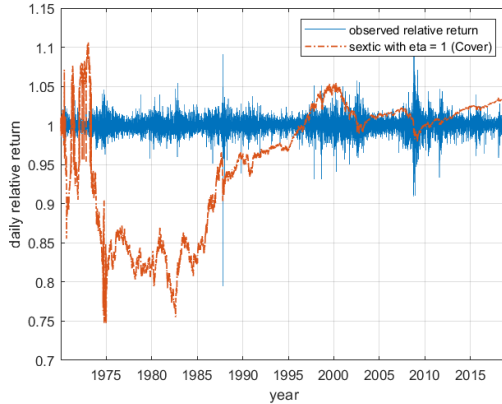
1. Propose a new prediction methodology for *relative stock index returns* based on universal portfolio construction
 - *relative stock index return* \equiv gross return of a stock index / (1 + interest rate)
 - An on-line strategy competitive with the best *constant proportional portfolio*
2. Derive a closed-form 6-order (sextic) predicting formula whose coefficients are **solely** determined by historical data
 - no other model assumptions or market assumptions

1. Propose a new prediction methodology for *relative stock index returns* based on universal portfolio construction
 - *relative stock index return* \equiv gross return of a stock index / (1 + interest rate)
 - An on-line strategy competitive with the best *constant proportional portfolio*
2. Derive a closed-form 6-order (sextic) predicting formula whose coefficients are **solely** determined by historical data
 - no other model assumptions or market assumptions
3. Empirically, the daily predictive errors in the 2010-2018 period is *below 2%*

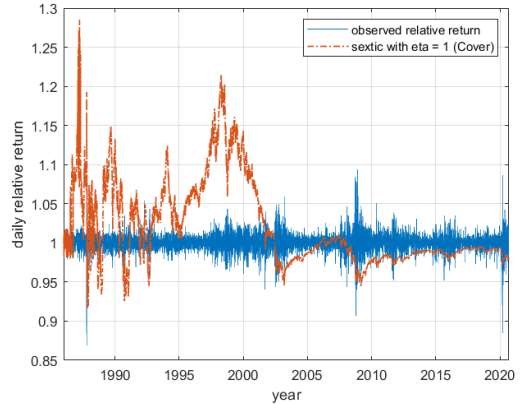
Market	Stock Index (Asset 2)	Risk Free Rate (Asset 1)	Sample Period	# of Obs (N)
US	S&P 500 index	12-month T-Bill	1970 - 2018	12,228
US	CRSP value weighted index	12-month T-Bill	1970 - 2018	12,228
UK	FTSE 100 index	12-month LIBOR	1986 - 2018	8,856
Japan	NIKKEI 225	12-month Government Bond	1981 - 2018	9,753
China	CSI 300 index	12-month deposit rate	2005 - 2018	3,265

- Use daily observations to ensure *sufficient* historical data

Relative Stock Index Return Prediction

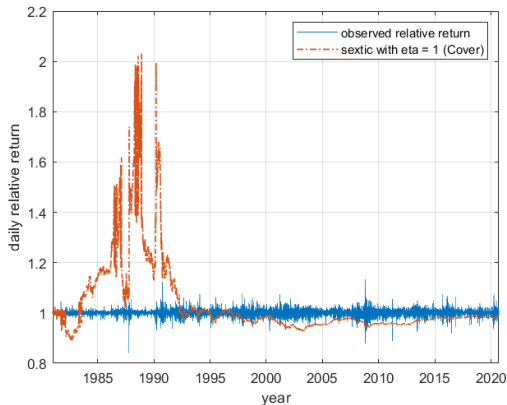


(a) S&P 500 vs 12M T-Bill (USA)

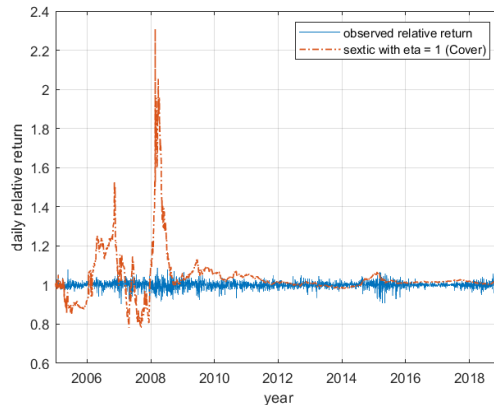


(b) FTSE 100 vs 12M LIBOR (UK)

Relative Stock Index Return Prediction



(c) NIKKEI 225 vs 12M gov bond (Japan)



(d) CSI 300 vs 12M deposit rate (China)

- An On-line machine learning prediction strategy for the decision maker
 - Deterministic prediction unconditionally? NO \Leftarrow Foster and Vohra (1999)
 - On-line machine learning? Promising \Leftarrow Cesa-Bianchi and Lugosi (2006)
 - To find the best on-line learning strategies for portfolio constructions against which the worst-case return data that **Nature** provides

- An On-line machine learning prediction strategy for the decision maker
 - Deterministic prediction unconditionally? NO \Leftarrow Foster and Vohra (1999)
 - On-line machine learning? Promising \Leftarrow Cesa-Bianchi and Lugosi (2006)
 - To find the best on-line learning strategies for portfolio constructions against which the worst-case return data that **Nature** provides
- Widely studied in Computer Science, Operational Research, Game Theory, etc.
 - Cover (1991), Cover and Thomas (1991), Hazan and Kale (2015), Hazan et al. (2007), Helmbold et al. (1998), and Vovk and Watkins (1998)

- To predict/forecast based entirely on **sufficient historical data**
- Without any assumptions on the market.
- Even though efficient numerical schemes can solve our problems, we are able to derive a closed-form solution that can be easily implemented

- Consider an economy of two assets with **sufficient** discrete periods, N
- In each single period, n ,
 - $x_n \equiv$ gross return of asset 1 (bond)
 - $y_n \equiv$ gross return of asset 2 (stock)
- A *relative stock return* is defined as,

$$R_n = \frac{y_n}{x_n}, \quad \Longleftrightarrow \quad r_n = \frac{1 - R_n}{R_n} = \frac{x_n}{y_n} - 1, \quad (1)$$

Constant proportional strategy I

- $b \in [0, 1]$: weight in asset 1 (bond) and remains constant through time.
- The cumulative return after N periods is,

$$S_N(b) \equiv \prod_{n=1}^N [bx_n + (1 - b)y_n] = \prod_{n=1}^N x_n \prod_{n=1}^N [b + (1 - b)R_n] \quad (2)$$

- R_n is the *relative stock return*

Constant proportional strategy II

- The best proportional strategy for the first N periods is thus

$$b_N^* = \operatorname{argmax}_{0 \leq b \leq 1} S_N(b). \quad (3)$$

which needs all information up to N (**unknown**).

Constant proportional strategy II

- The best proportional strategy for the first N periods is thus

$$b_N^* = \operatorname{argmax}_{0 \leq b \leq 1} S_N(b). \quad (3)$$

which needs all information up to N (**unknown**).

- Logarithm transformation and first-order condition (FOC)

$$\sum_{n=1}^N \frac{1 - R_n}{R_n + (1 - R_n)b_N^*} = 0 \quad \Longleftrightarrow \quad \sum_{n=1}^N \frac{r_n}{1 + r_nb_N^*} = 0. \quad (4)$$

- An on-line strategy by **Thomas M. Cover (Mathematical Finance, 1991)**
 - μ : a distribution on $[0, 1]$, either uniform or Dirichlet $(\frac{1}{2}, \frac{1}{2})$ distributed.
 - \hat{b}_n : the weight of asset 1 (risk-free bond),

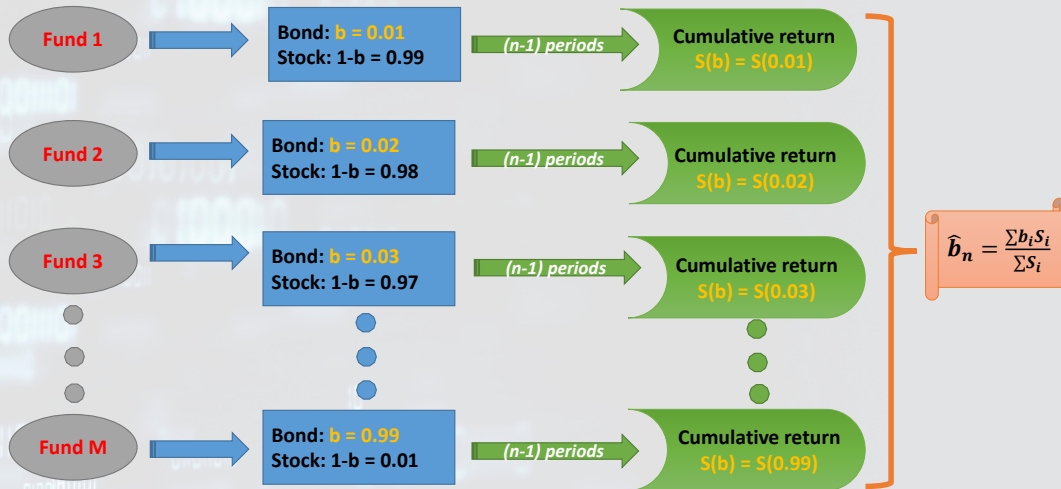
$$\hat{b}_1 = \frac{1}{2}; \quad \hat{b}_n = \frac{\int_0^1 b S_{n-1}(b) d\mu(b)}{\int_0^1 S_{n-1}(b) d\mu(b)}, \quad n \geq 2 \quad (5)$$

- \hat{b}_n is conditionally known at the time $(n - 1)$

- Following this strategy, the cumulative return for the first N periods is,

$$\hat{S}_N \equiv \prod_{n=1}^N [\hat{b}_n x_n + (1 - \hat{b}_n) y_n] = \prod_{n=1}^N x_n \prod_{n=1}^N [\hat{b}_n + (1 - \hat{b}_n) R_n]. \quad (6)$$

Intuition – Fund of Funds



Algorithm 1 Pseudo algorithm for Cover (1991) on-line strategy

1: Initialization: $b = [0.01 : 0.01 : 0.99]^T$

2: **for** $n = 2$ to N **do**

3: **for** $i = 1$ to $\text{length}(b)$ **do**

4: $b_i \leftarrow b(i)$

5: $S_i \leftarrow \prod_{t=1}^{n-1} [b_i x_t + (1 - b_i) y_t]$

▷ cumulative return of $N - 1$ periods given b_i

6: $\omega_i \leftarrow \frac{S_i}{\sum_{i=1} S_i}$

▷ the contribution of i -th strategy b_i

7: $\hat{b}_n \leftarrow \sum_i \omega_i * b_i$

▷ the on-line strategy for the period n

A Rough Comparison

Optimal constant strategy

$$S_N^* = \prod_{n=1}^N x_n \prod_{n=1}^N [b_N^* + (1 - b_N^*)R_n]$$

- $b_N^* = \operatorname{argmax}_{0 \leq b \leq 1} S_N(b)$
- needs the information in all N time periods

On-line ML strategy

$$\hat{S}_N = \prod_{n=1}^N x_n \prod_{n=1}^N [\hat{b}_n + (1 - \hat{b}_n)R_n]$$

- $\hat{b}_n = \frac{\int_0^1 b S_{n-1}(b) d\mu(b)}{\int_0^1 S_{n-1}(b) d\mu(b)}$
- needs the information in $(N - 1)$ time periods

- Cover (1991) proves that

$$1 \leq \frac{S_N^*}{\hat{S}_N} \leq 2\sqrt{N+1}. \quad (7)$$

- A universal portfolio is the on-line strategy with the property such that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \left(\frac{S_N^*}{\hat{S}_N} \right) = 0. \quad (8)$$

- Rewrite Equation (7) with a constant, $c_N \in [0, 1]$

$$\log \left(\frac{S_N^*}{\hat{S}_N} \right) = \log(2c_N) + \frac{\log(N+1)}{2}, \quad 0 \leq c_N \leq 1 \quad (9)$$

- Drop $\log(2c_N)$ and express the left side above with r_n such that,

$$r_n = \frac{1 - R_n}{R_n} = \frac{x_n}{y_n} - 1$$

Solve $\{b_N^*, \hat{r}_N\}$ Simultaneously

1. Function from universal portfolio construction

$$\left[\log(1 + b_N^* \hat{r}_N) - \log(1 + \hat{b}_N \hat{r}_N) \right] + \sum_{n=1}^{N-1} \left[\log(1 + b_N^* r_n) - \log(1 + \hat{b}_n r_n) \right] = \frac{\log(N+1)}{2}. \quad (10)$$

Solve $\{b_N^*, \hat{r}_N\}$ Simultaneously

1. Function from universal portfolio construction

$$\left[\log(1 + b_N^* \hat{r}_N) - \log(1 + \hat{b}_N \hat{r}_N) \right] + \sum_{n=1}^{N-1} \left[\log(1 + b_N^* r_n) - \log(1 + \hat{b}_n r_n) \right] = \frac{\log(N+1)}{2}. \quad (10)$$

2. Function from first-order condition

$$\frac{\hat{r}_N}{1 + \hat{r}_N b_N^*} + \sum_{n=1}^{N-1} \frac{r_n}{1 + r_n b_N^*} = 0. \quad (11)$$

How Close between r_N and \hat{r}_N ?

Proposition 1

- Assume that $0 < m \leq r_n \leq M, \forall n$ for two fixed positive numbers m and M .
- By viewing b_N^* as a function of $\{r_1, \dots, r_N\}$, its partial derivative to r_N is uniformly bounded from below by a positive constant c .
- Then the solution of the Equation (10) and (11), \hat{r}_N , approximates to r_N as close as possible when N is large enough.

How Close between r_N and \hat{r}_N ?

- As a function of the data $\{r_1, \dots, r_N\}$, we can show that,

$$\frac{\partial b_N^*}{\partial r_N} > 0. \quad (12)$$

- When the relative return of the asset 1 to the asset 2 is increased, the proportion on the asset 1 is increased as well following the strategy b_N^*

How Close between r_N and \hat{r}_N ?

- Proposition 1 states that if the contribution of r_N to b_N^* is *non-degenerate* in the sense that, for a positive number c ,

$$\frac{\partial b_N^*}{\partial r_N} \geq c, N = 1, 2, \dots, \quad (13)$$

- then

$$r_N - \hat{r}_N = o(1). \quad (14)$$

- Introduce the following notations by historical data only

$$P_1 = \sum_{n=1}^{N-1} r_n, \quad P_2 = \sum_{n=1}^{N-1} r_n^2, \quad Q_1 = \sum_{n=1}^{N-1} \hat{b}_n r_n, \quad Q_2 = \sum_{n=1}^{N-1} \hat{b}_n^2 r_n^2.$$

- To derive a analytical formula for \hat{r}_N with Taylor expansion
 - 2nd-order for $\log(x) \implies$ sextic formula
 - 1st-order for $\log(x) \implies$ cubic formula

A Sextic Predicting Formula

- \hat{r}_N can be solved from the following formula,

$$\begin{aligned} f_6(x) \equiv & \hat{b}_N^2 x^6 - 2\hat{b}_N x^5 + \left[Q_2 - 2Q_1 + 2P_2 \hat{b}_N - \log(N+1) + 1 \right] x^4 \\ & + (2P_1 - 2P_2 \hat{b}_N) x^3 + \left[P_1^2 - 2P_2 Q_1 + P_2 + P_2^2 \hat{b}_N^2 - 2P_2^2 \log(N+1) \right] x^2 \\ & + (2P_1 P_2 - 2P_2^2 \hat{b}_N) x + \left[P_1^2 P_2 - 2P_2^2 Q_1 + P_2^2 Q_2 - P_2^2 \log(N+1) \right], \end{aligned}$$

- whose coefficients are **solely** determined by historical data.

Alternative Construction of \hat{b}_n

- Vovk and Watkins (1998)

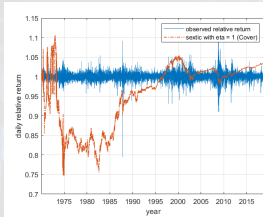
$$\hat{b}_n^{VW} = \frac{\int_0^1 b S_{n-1}(b)^\eta db}{\int_0^1 S_{n-1}^\eta db}, \quad 0 < \eta \leq 1, n \geq 2 \quad (15)$$

- When $\eta = 1$,

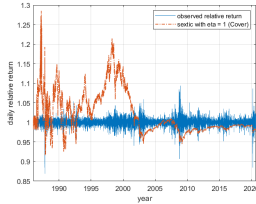
$$\hat{b}_n^{VW} \implies \hat{b}_n$$

Predicted Relative Stock Index Return

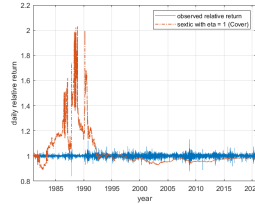
- red line: sextic formula predicted *relative stock index return*
- blue line: market observed *relative stock index return*



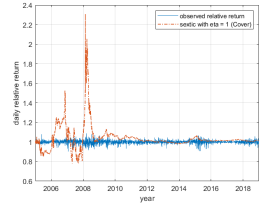
(e) S&P 500 vs 12M T-Bill
(USA)



(f) FTSE 100 vs 12M LIBOR
(UK)



(g) NIKKEI 225 vs 12M gov
bond (Japan)



(h) CSI 300 vs 12M deposit
rate (China)

- We evaluate the prediction performance during the period 2010 - 2018

$$\text{predictive error} \equiv \frac{\hat{R}_N - R_N}{R_N}. \quad (16)$$

- Consider 4 statistics for the time series of predictive errors suggested by Cremers (2002)
 1. *MAD*: the average of the absolute predictive errors
 2. *Bias*: the average of the predictive errors
 3. *Deviation*: the average of the squared difference between the bias and the predictive errors
 4. *RMSE*: the square root of the average of the sum of squared predictive errors.

S&P 500 index vs 12-month T-Bill

$$\hat{b}_1^{VW} = 0.5, \quad \hat{b}_n^{VW} = \frac{\int_0^1 b S_{n-1}(b)^\eta db}{\int_0^1 S_{n-1}^\eta db}, \quad 0 < \eta \leq 1, n \geq 2$$

	$\eta = 1$ (Cover)	$\eta = 0.75$	$\eta = 0.5$	$\eta = 0.15$	$\eta = 0.05$
MAD	1.96%	2.00%	2.06%	2.13%	2.15%
Bias	1.78%	1.83%	1.88%	1.96%	1.98%
Deviation	0.02%	0.02%	0.02%	0.02%	0.02%
RMSE	2.29%	2.34%	2.40%	2.48%	2.51%

CRSP value index vs 12-month T-Bill

$$\hat{b}_1^{VW} = 0.5, \quad \hat{b}_n^{VW} = \frac{\int_0^1 b S_{n-1}(b)^\eta db}{\int_0^1 S_{n-1}^\eta db}, \quad 0 < \eta \leq 1, n \geq 2$$

	$\eta = 1$ (Cover)	$\eta = 0.75$	$\eta = 0.5$	$\eta = 0.15$	$\eta = 0.05$
MAD	1.14%	1.16%	1.18%	1.20%	1.21%
Bias	0.23%	0.24%	0.25%	0.26%	0.26%
Deviation	0.02%	0.02%	0.02%	0.02%	0.02%
RMSE	1.42%	1.44%	1.46%	1.49%	1.50%

FTSE 100 index vs 12-month LIBOR

$$\hat{b}_1^{VW} = 0.5, \quad \hat{b}_n^{VW} = \frac{\int_0^1 b S_{n-1}(b)^\eta db}{\int_0^1 S_{n-1}^\eta db}, \quad 0 < \eta \leq 1, n \geq 2$$

	$\eta = 1$ (Cover)	$\eta = 0.75$	$\eta = 0.5$	$\eta = 0.15$	$\eta = 0.05$
MAD	1.46%	1.48%	1.51%	1.54%	1.54%
Bias	-1.35%	-1.37%	-1.40%	-1.43%	-1.44%
Deviation	0.01%	0.01%	0.01%	0.01%	0.01%
RMSE	1.80%	1.82%	1.84%	1.87%	1.88%

NIKKEI 225 index vs 12-month Gov-bonds

$$\hat{b}_1^{VW} = 0.5, \quad \hat{b}_n^{VW} = \frac{\int_0^1 b S_{n-1}(b)^\eta db}{\int_0^1 S_{n-1}^\eta db}, \quad 0 < \eta \leq 1, n \geq 2$$

	$\eta = 1$ (Cover)	$\eta = 0.75$	$\eta = 0.5$	$\eta = 0.15$	$\eta = 0.05$
MAD	2.81%	2.89%	2.97%	3.10%	3.13%
Bias	-2.70%	-2.79%	-2.88%	-3.01%	-3.04%
Deviation	0.03%	0.03%	0.03%	0.03%	0.03%
RMSE	3.23%	3.31%	3.39%	3.52%	3.55%

CSI 300 index vs 12-month Deposit Rate

$$\hat{b}_1^{VW} = 0.5, \quad \hat{b}_n^{VW} = \frac{\int_0^1 b S_{n-1}(b)^\eta db}{\int_0^1 S_{n-1}^\eta db}, \quad 0 < \eta \leq 1, n \geq 2$$

	$\eta = 1$ (Cover)	$\eta = 0.75$	$\eta = 0.5$	$\eta = 0.15$	$\eta = 0.05$
MAD	4.27%	4.33%	4.40%	4.49%	4.52%
Bias	4.21%	4.28%	4.35%	4.45%	4.47%
Deviation	0.11%	0.11%	0.11%	0.11%	0.12%
RMSE	5.33%	5.40%	5.48%	5.59%	5.62%

- Economic meanings behind the on-line machine learning strategy
 - Two time series: any similarities between them? Patterns?
- Option strategy based on the predicted relative return
 - If we have available (sufficient) option data, we can construct an option trading strategy based on our 2% error bounded predicted value

- Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge university press.
- Cover, T. M. (1991). Universal portfolios. *Mathematical Finance*, 1(1):1–29.
- Cover, T. M. and Thomas, J. A. (1991). Elements of information theory. *New York: Wiley*, 3:37–38.
- Cremers, K. M. (2002). Stock return predictability: A bayesian model selection perspective. *The Review of Financial Studies*, 15(4):1223–1249.
- Foster, D. P. and Vohra, R. (1999). Regret in the on-line decision problem. *Games and Economic Behavior*, 29(1-2):7–35.
- Hazan, E., Agarwal, A., and Kale, S. (2007). Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192.
- Hazan, E. and Kale, S. (2015). An online portfolio selection algorithm with regret logarithmic in price variation. *Mathematical Finance*, 25(2):288–310.
- Helmbold, D. P., Schapire, R. E., Singer, Y., and Warmuth, M. K. (1998). On-line portfolio selection using multiplicative updates. *Mathematical Finance*, 8(4):325–347.
- Vovk, V. and Watkins, C. (1998). Universal portfolio selection. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 12–23.