Predict Relative Stock Index Returns by Sufficient Historical Data

Abstract

This paper presents a new prediction methodology on relative stock index returns - the ratio of the stock index return to the interest rate. This prediction methodology relies on the universal portfolio construction and historical data alone. Following this machine learning approach, the predictive errors for the daily prediction of the relative stock index return in the 2010-2018 period can be as small as 2 percent. We find out that the relative predictive errors depend on the number of historical data and the empirical covariance structure as the theory predicts. This paper presents a surprising application of machine learning to the stock return prediction problem, and this new prediction methodology is promising by our empirical result.

Keywords: universal portfolio, on-line machine learning, return predictability

JEL Classification Codes: G01, G12, G14, G20.

1 Introduction

While most finance people believe that the stock returns are hardly predictable, in this paper, we demonstrate that the ratio of the stock return to the interest rate in the same period, *relative stock index return*, is **predictable asymptotically**. This new prediction methodology relies entirely on sufficient historical data without using any assumption on the market.

If one wants to find a *deterministic* prediction of the stock return unconditionally, the answer is simply no. The reason is straightforward. Whatever a deterministic prediction formula is proposed, Nature (the adversary) can come up a data with the largest possible error with the proposed scheme (see the famous arguments for what is a well-defined prediction problem in Foster and Vohra, 1999). There are two general learning approaches to a well-defined prediction problem, and each of them has been greatly successful. One is to impose a stochastic assumption about the market factor (statistical learning approach, see Hastie, Tibshirani and Friedman, 2016). Another is on-line machine learning approach to study a mixed (probabilistic) strategy for the decision marker instead of a deterministic strategy (Cesa-Bianchi and Lugosi, 2006). This paper is closely related to the on-line machine learning approach to the prediction problem.

Specifically, this paper provides a new application of the on-line machine learning framework to the prediction problem of the relative stock return. This proposed prediction methodology depends on a strand of researches on universal portfolio construction. Motivated by information theory, a universal portfolio is one on-line strategy that is competitive with the best constant proportional portfolio in highlight. This class of on-line strategy has been well studied in computer science, operational research, and economics (see, for instance, Cover and Thomas, 1991; Cesa-Bianchi and Lugosi, 2006). Many recent researchers aim to find the best on-line learning strategies for the portfolio construction against which Nature provides the worst-case return data. For instance, Cover (1991), Hazan and Kale (2015), Hazan, Agarwal and Kale (2007), Helmbold, Schapire, Singer and Warmuth (1998), and Vovk and Watkins (1998), among others, have proposed several efficient algorithms to construct the universal portfolio. Distinguished from these previous studies, the universal portfolio is an *input*, not an output, in our approach to the relative stock return prediction.

By using the universal portfolio construction, we derive a system of deterministic non-linear equations of the relative stock index return in the future. All coefficients of these equations are computed by historical (available) data. Surprisingly, we demonstrate that in an asymptotic sense the real stock return can be solved from these non-linear equations as close as possible. Our proof is based on ingenious theorems in the universal portfolio literature.

This prediction methodology can be implemented easily. In particular, we derive a sextic polynomial prediction formula for the relative stock index return. According to our empirical results, this prediction method is promising. We find out that the more historical data, the more accurate the prediction in general. The variation of the historical relative stock return also plays a crucial role in prediction: the smaller variation of the historical data, the better the prediction. For instance, using historical data alone to predict the relative stock index return in the 2010-2018 period, the relative predictive errors for the daily prediction can be as small as 2 percent. Furthermore, the data quality also affects the prediction outcome. Overall, this prediction methodology is supportive by our empirical results.

In Section 2, we explain the prediction methodology and introduce three analytical prediction formulae, including the sextic polynomial prediction formula. We present the data and empirical results in Section 3 and Section 4, respectively. Conclusion is given in Section 5 and Appendix A contains a technical part of the prediction methodology.

2 Prediction Methodology

We consider an economy of two assets with sufficiently many discrete periods. Briefly speaking, asset 1 represents the fixed income market and asset 2 represents the equity market. In each single time period, n, let x_n and y_n denote the gross holding period return of asset 1 and 2, respectively.We assume that $x_n > 0$ and $y_n > 0$ for all possible data. Define,

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

R_n is the relative asset return (asset 2 versus asset 1), and R_n and r_n are determined from each other.

Given **sufficient** historical data, $\{x_n, y_n, 1 \leq n \leq N-1\}$, can we predict the relative asset return R_N to certain degree? In this section, we present a new prediction methodology for the relative stock index return and several analytical prediction formulas.

2.1 A Benchmark Formula

We start with a simple prediction formula that is based on classical asset pricing theory. Due to its simple assumption, this prediction method merely serves as a benchmark for subsequent discussions.

To highlight the role of historical data (i.e., the number of periods), we use $\mathcal{N}$ to represent a market with N investment periods. There is one representative agent "N" and the objective of this investor is to maximize its cumulative growth return. In other words, this investor's preference is a log utility that maximizes the expected growth rate.

Since we do not impose any model assumptions on x_n or y_n, we focus on the class of constant proportional strategies, the Kelly's rule.[1] Specifically, Let $b \in [0, 1]$ denote the wealth proportion allocated in asset 1 and it remains constant through time. The cumulative return after N periods is,

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

Therefore, the best proportional strategy for the first N periods, $b_N^\star$, is determined by

$$b_N^\star = \underset{0 \le b \le 1}{\operatorname{argmax}}\ S_N(b).$$

Taking the logarithm transformation and applying the first-order condition, we obtain

$$\sum_{n=1}^{N} \frac{1 - R_n}{R_n + (1 - R_n)b_N^\star} = 0. \tag{1}$$

In a pure exchange equilibrium model (see, for instance, Cochrane, 2000), the representative agent should invest all in the risky asset (asset 2) and zero in risk-free asset (asset 1) as the bond is in zero supply in equilibrium. Therefore, $b_N^\star$ must be zero in Equation (1), implying that,

$$\frac{1}{R_N} = N - \sum_{n=1}^{N-1} \frac{1}{R_n}. \tag{2}$$

We cite Equation (2) as the benchmark (or standard equilibrium) prediction using historical data, but we notice that many complicated equilibrium models have been studied in literature under further model or market assumptions.

2.2 Prediction from the universal portfolio

In this section, we introduce a prediction methodology without relying on the equilibrium assumption on the market or model assumptions about future assets' returns. Rather, we use the

[1] It is well known that, under certain assumptions about the market such as a stationary market, the Kelly's rule is equivalent to the maximization of the expected log-utility in Merton's standard portfolio choice framework. See Algoget (1992), Bremain (1961).

universal portfolio motivated by information theory (Cover, 1991). Let μ denote a distribution on $[0,1]$ (either uniform or Dirchlet $(\frac{1}{2},\frac{1}{2})$ distributed). We introduce the following on-line strategy at each time period n. Let $\hat{b}_n$ denote the weight invested in asset 1,

- $\hat{b}_1 = \frac{1}{2}$, an equally-weighted investment strategy at the first time period.
- For each $n \geq 2$,

$$\hat{b}_n = \frac{\int_0^1 bS_{n-1}(b)d\mu(b)}{\int_0^1 S_{n-1}(b)d\mu(b)}. \tag{3}$$

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

$$\hat{S}_N = \prod_{n=1}^{N}\left(\hat{b}_n x_n + (1-\hat{b}_n)y_n\right) = \prod_{n=1}^{N} x_n \prod_{n=1}^{N}\left(\hat{b}_n + (1-\hat{b}_n)R_n\right).$$

When the representative investor searches the best strategy $b_N^\star$ in the first N time periods, he actually needs the information in all N time periods. In other words, he must know the future market $\{b_1, \cdots, b_N\}$ in order to find the best strategy $b_N^\star$. On the contrary, the strategy $\hat{b}_n$ depends only on the historical data $\{R_1, \cdots, R_{n-1}\}$ up to time period $n-1$, for any time n. The strategy that depends only available historical data is called *on-line* strategy in the machine learning literature (see, for instance, Hazan, Agarwal and Kale, 2007).

For any $N \geq 1$, let $S_N^\star = \max_{0\leq b\leq 1} S_N(b)$. Cover (1991) shows that

$$1 \leq \frac{S_N^\star}{\hat{S}_N} \leq 2\sqrt{N+1}. \tag{4}$$

As a consequence,[2]

$$\lim_{N\to\infty}\frac{1}{N}\log\left(\frac{S_N^\star}{\hat{S}_N}\right) = 0. \tag{5}$$

We rewrite Equation (4) as

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

By expressing the left side of the last equation in terms of r_n, we have,

$$\sum_{n=1}^{N}\left[\log(1+b_N^\star r_n) - \log(1+\hat{b}_n r_n)\right] = \log(2c_N) + \frac{\log(N+1)}{2}. \tag{6}$$

where $b_N^\star$ satisfies Equation (1). That is,

$$\sum_{n=1}^{N}\frac{r_n}{1+r_n b_N^\star} = 0. \tag{7}$$

[2]The on-line strategy with this property is called a universal portfolio.

Our idea is to solve $\{R_N, b_N^\star\}$ simultaneously by using Equations (6) and (7). However, Equation (6) is not well-defined in the presence of the unknown constant c_N. Our prediction methodology is to replace Equations (6) by the following equation

$$\sum_{n=1}^{N}\Big[\log(1+b_N^\star r_n)-\log(1+\hat{b}_n r_n)\Big]=\frac{\log(N+1)}{2}. \tag{8}$$

The solution of Equation (7) and (8) is guaranteed to approximate r_N when the number of the market data is large enough, by the following result.

Proposition 2.1. *Assume that $0<m\le r_n\le M, \forall n$ for two fixed positive numbers m and M. Moreover, by viewing b_N^* as a function of $r_1,\cdots,r_N$, its partial derivative to r_N is uniformly bounded from below by a positive constant c, then the solution of the Equation (7) and (8) approximates to r_N as close as possible when N is large enough.*

Proof. See Appendix A. □

In this proposition, the first assumption about r_n is reasonable because both x_n and y_n are bounded in the market data. As a function of the data $\{r_1,\cdots,r_N\}$, we can show that (See Lemma 5.1 in Appendix A)

$$\frac{\partial b_N^*}{\partial r_N}>0. \tag{9}$$

Its intuition is simple. When the relative return of the asset 1 to the asset 2 is increased, the proportion on the asset 1 is increased as well by following the strategy b_N^*. Proposition 2.1 states that if the contribution of r_N to b_N^* is *non-degenerate* in the sense that, for a positive number c,

$$\frac{\partial b_N^*}{\partial r_N}\ge c, N=1,2,\cdots, \tag{10}$$

then [3]

$$r_N-\hat{r}_N=o(1) \tag{11}$$

where $\hat{r}_N$ is one solution about r_N in the linear systems (8) and (7).

What if $\frac{\partial b_N^*}{\partial r_N}$ is degenerate? It means that $\frac{\partial b_N^*}{\partial r_N}\sim 0$; then the market data r_N at time N has no significant effect to the best constant proportion strategy b_N^*. Equivalently, the historical data $r_1,\cdots,r_{N-1}$ are virtually enough to find b_N^*, a prediction can be done. Therefore, from the prediction perspective, it is also reasonable to impose such a non-degenerate condition in (10).

[3]By $a_N=o(1)$ we means that $a_N\to 0$ when $N\to 0$.

There are two unknown variables b_N^* and r_N in Eq. (7) and (8) given the historical data $\{R_1, \cdots, R_{N-1}\}$. By its definition, $\log S_N^\star$ is a function of $\{R_1, \cdots, R_{N-1}, R_N\}$. On the other hand, $\hat{b}_n$ depends only on $\{R_1, \cdots, R_{n-1}\}$, as shown in Equation (3). $\sum_{n=1}^{N} \log(1 + \hat{b}_n r_n)$ is thus a function of R_N. Therefore, Equation (8) becomes one equation of R_N in terms of historical data and $b_N^\star$. The prediction methodology is to solve Equations (7) and (8) *deterministically* of the two variables $\{R_N, b^*\}$.

2.3 Prediction Formulas

Even though efficient numerical schemes can solve equations (7) and (8), we intend to derive an analytical prediction formula of r_N in this subsection to highlight this prediction approach. From a technical point of view, the roots in Equations (7) and (8) are *continuous* to the coefficients in these two equations; thus, are continuous to the available data. Therefore, in what follows, we approximate the non-linear functions, $\log(1 + x)$ and $\frac{1}{1+x}$, to derive **analytical prediction** formulas.

For simplicity, we introduce the following notations.

$$P_1 = \sum_{n=1}^{N-1} r_n, \qquad P_2 = \sum_{n=1}^{N-1} r_n^2,$$

and

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

Both P_1 and P_2 simply depend on historical data. Q_1 and Q_2 also depend on the historical data. However, how to compute $\hat{b}_n$ using historical data is also important to derive Q_1 and Q_2. The most straightforward way to construct $\hat{b}_n$ is to adopt Cover's rule in Eq. (3). In the next section, we introduce alternative constructions for $\hat{b}_n$.

We obtain two prediction formulae for r_N, being a solution of a cubic polynomial and a sextic polynomial in which all coefficients are entirely determined by historical data. As will be shown later, the sextic degree polynomial prediction formula works much better than the cubic one; thus, the cubic prediction one is often used to explain the approximation value of the sextic polynomial prediction formula.

2.3.1 A cubic polynomial prediction formula

Let

$$f_3(x) \equiv \hat{b}_N x^3 + \left[Q_1 - 1 + \frac{\log(N+1)}{2}\right] x^2 + \left(P_2\hat{b}_N - 2P_1\right) x + \left[P_2\frac{\log(N+1)}{2} - P_1^2 + P_2Q_1\right].$$

All coefficients of the cubic polynomial $f_3(x)$ are computed by the historical data. For the cubic equation $f_3(x) = 0$, there exists at least one real root. We choose the real root, r_N^{FF}, that is closest to r_{N-1}. r_N^{FF} is used to predict r_N in the next time period, N.

2.3.2 A sextic polynomial prediction formula

Define a sextic polynomial function

$$\begin{aligned} f_6(x) &= \hat{b}_N^2 x^6 - 2\hat{b}_N x^5 + \left[Q_2 - 2Q_1 + 2P_2\hat{b}_N - \log(N+1) + 1\right] x^4 + (2P_1 - 2P_2\hat{b}_N)x^3 \\ &+ \left[P_1^2 - 2P_2Q_1 + P_2 + P_2^2\hat{b}_N^2 - 2P_2^2\log(N+1)\right] x^2 + (2P_1P_2 - 2P_2^2\hat{b}_N)x \\ &+ \left[P_1^2P_2 - 2P_2^2Q_1 + P_2^2Q_2 - P_2^2\log(N+1)\right] \end{aligned}$$

All coefficients of $f_6(x)$ are derived from the historical data; however, the existence of a real root is not guaranteed for a general sextic equation. We choose the real number, r_N^{SF}, that is closest to r_{N-1} and meanwhile $f_6(r_N^{SF})$ is close to zero (See Appendix A for its construction).

2.4 Alternative constructions of $\hat{b}_N$

In the above two prediction formulas, an explicit construction of the on-line strategy $\hat{b}_n$ is crucial. The construction of the on-line strategy $\hat{b}_n$ is far from unique with similar properties in Cover's theorem. For instance, for a positive number $\eta \in (0, 1]$, let

$$\hat{b}_{n+1}^{EG} = \hat{b}_n^{EG}\frac{exp(\eta p_n)}{exp(\eta p_n) + exp(\eta q_n)}, \tag{12}$$

where

$$p_n \equiv \frac{\hat{b}_n^{EG}x_n}{\hat{b}_n^{EG}x_n + (1-\hat{b}_n^{EG})y_n} = \frac{\hat{b}_n^{EG}}{\hat{b}_n^{EG} + (1-\hat{b}_n^{EG})R_n},$$

and

$$q_n \equiv \frac{(1-\hat{b}_n^{EG})R_n}{\hat{b}_n^{EG} + (1-\hat{b}_n^{EG})R_n}.$$

Helmbold, Schapire, Singer and Warmuth (1998) demonstrate that the $\log\left(\frac{S_N^*}{\hat{S}_N^{EG}}\right)$ is bounded by $c\sqrt{N}$ for a constant c.

As another example, $\hat{b}$ is constructed such that,

$$\hat{b}_{n+1}^{VW} = \frac{\int_0^1 bS_{n-1}(b)^\eta db}{\int_0^1 S_{n-1}^\eta db}. \tag{13}$$

Vovk and Watkins (1998) show that $\log\left(S_N^*/\hat{S}_N^{VW}\right)$ is bounded by $c + \frac{log(N+1)}{2\eta}$.

Compared with Cover's original construction of $\hat{b}$, two alternatives, $\hat{b}^{EG}$ and $\hat{b}^{VW}$, share a common feature that a higher allocation weight is placed on the asset with better performance in the previous period. In our empirical testing below, we report prediction results using $\hat{b}^{VW}$ as the main insights from other on-line strategy constructions are fairly similar.

3 Data Description

From this section, we report our empirical results of the proposed prediction methodology. To be robust, we use both monthly return and daily return in each period to document the effect of the number of period N.

We use asset 1 to represent the bond market and its return is proxied by 3-month or 1-year T-Bill rate for the following reasons. On one hand, since the earliest T-bill reported for monthly data is 3-month T-Bill, we use 3-month T-Bill as a proxy for asset 1 to convert and construct *monthly* return of asset 1 and we obtain T-Bill rate from Federal Reserve Bank of St. Louis. On the other hand, when we consider the daily return to increase the number of period, there is no data reported for 3-month T-Bill until January, 1982. Thus, we use 1-year T-Bill as a proxy for asset 1 to construct *daily* return instead.

The asset 2, representing the equity market, is proxied by S&P 500 index, CRSP value weighted index, or Wilshire 5000 index. We obtain the historical data of S&P 500 index, CRSP value weighted index and Wilshire 5000 index from the Center for Research in Security Prices (CRSP). Specifically, the relative return is calculated as, $R_n = \frac{y_n}{x_n}$, where x_n is 1 plus T-bill rate and y_n is 1 plus the net return of either S&P 500 index, CRSP value index or Wilshire 5000 index in the same period.

As for the data sample, for S&P 500 index and CRSP value weighted index, our monthly sample ranges from January, 1934 to December, 2018 with a sample size of 1020 and the daily sample from January, 1970 to December, 2018 with a sample size of 12228. In contrast, the monthly sample for Wilshire 5000 ranges from February, 1971 to December 2018 and the daily sample from June, 1981 to December 2018. Due to the poor data quality, we remove the earlier years where Wilshire 5000 index levels are reported as zero. We plot the monthly and daily relative return of S&P 500 to T-Bill in Figure 1 (a)-(b), respectively.

4 Empirical Analysis

In this section, we present the empirical results. From a theoretical perspective, the performance of the predicting formulas depends on both the number of historical data and the empirical covariance structure. Indeed, the longer the data period, the more accurate the prediction because of Equation (4) and Proposition 2.1. Moreover, the financial market in the real world is high volatile and the empirical volatility affects the performance of universal portfolio (See, for instance, Hazan and Kale, 2015). Therefore, we perform the prediction exercise through different samples of historical data to include a series of financial crises and extreme events in the financial market. To be consistent, we focus on evaluating the predictive performance of the most recent decade in the **period of 2010 - 2018**.

We define the predictive error as,[4]

$$\text{predictive error} \equiv \frac{\hat{R}_n - R_n^M}{R_n^M}, \tag{14}$$

where $\hat{R}_n$ is the predicted relative return using one of predicting formulas and R_n^M is the market observed relative return at time n.

After obtaining a time-series of forecast errors for each predicting formula, we calculate four different statistics, suggested by Cremers (2002), to evaluate the prediction performance: *MAD*, the average of the absolute predictive errors; *Bias*, the average of the predictive errors; *Deviation*, the average of the squared difference between the bias and the predictive errors; *RMSE*: the square root of the average of the sum of squared predictive errors.

4.1 Prediction in S&P 500 index

We first discuss the prediction by the monthly S&P 500 index data. Figure 2 compares the three predicting formulas. The cubic formula provides several extremely large deviations. As for the standard benchmark, after some significant fluctuations between 1940 and 1950, it consistently underestimate the predicted return after 1950, though the deviation magnitude is much smaller than that of cubic formula.

[4]Another error measure is simply, predictive error $\equiv \hat{R}_n - R_n^M$. We repeat our analysis with this error and the outcomes are consistent with our main results. The results are upon request.

As expected, the sextic polynomial formula gradually returns back to the real level as we use more and more historical data, in spite of some apparent overreactions during the period of 1950 to 1970. From the data variation perspective, the fluctuation during the period of 1930-1970 is not a surprise as financial market has experienced some severe challenges, such as the Great Depression in 1930s, the WWII in 1940s, and the energy crisis in 1970s, to name a few. The sextic polynomial formula dominates the cubic formula and the benchmark from standard equilibrium.

The alternative construction of $\hat{b}_n$ from Equation (13) does not yield big improvement. For instance, there is no significant difference by choosing different values of η, compared with the original construction of $\eta = 1$ (Cover's formula). In fact, the prediction demonstrated in Figure 2(d) with $\eta = 0.5$ is almost the same as that in Figure 2(c) with $\eta = 1$.

To demonstrate the effect of a large number of market data, Figure 3 compares the three predicting formulas with respect to "daily" relative return of S&P 500 index to 1-year T-bill. Consistent with what we have observed before, the sextic polynomial formula outperforms others as we use more and more historical data. The predicted relative return error using the sextic polynomial formula is at most 5 percent after the year of 1990. In the same time period, the predicted relative return error is bounded by 20 percent by using the cubic formula performs. The standard benchmark becomes the worst method as the data frequency increases.

Apparently, the sextic polynomial formula dominates other methods in both monthly return prediction and daily return prediction. And the predicting performance becomes better as we have sufficient historical data.

4.2 Performance Evaluation

We next examine the performance evaluation more closely by looking at some statistic measures.

We first consider the monthly relative return of S&P 500 index. Table 1 reports various predictive statistics during the period of 2010 - 2018 using monthly relative return of S&P 500 index to 3-month T-Bill. Each column corresponds to a different prediction formula and each panel uses the different sample prediction to initialize the prediction. In Panel A, we use full sample starting from 1934 to start the prediction. Consistent with Figure 2, both the cubic formula and the benchmark formula are not very reliable. For instance, the largest *MAD* is 531.90% by using the cubic formula. In sharp contrast, the sextic polynomial formula performs much better with the lowest *MAD*, *Bias*, *Deviation*, and *RMSE*. When $\eta = 1$, the *MAD* is around 7.36% , and the *RMSE* is around 8.30%.

Also, as we have seen in the time-series plots, a different construction of $\hat{b}_n$ with different parameter η does not yield obvious changes to the statistics of relative errors. Actually, the original construction of $\hat{b}_n$ with $\eta = 1$ is still the best choice, in terms of all four error statistics.[5]

We use the different sample period to start the prediction in Panel B to D. Within each panel, the sextic polynomial formula dominates other methods. Cover's original construction of $\hat{b}_n$ with $\eta = 1$ outperforms other choices. Comparing across each panel, the *MAD* and *Bias* for the sextic polynomial formulas are bounded below 10% and 13%, respectively. Surprisingly, we find that using the subsample period of 1970 - 2018 yields the lowest *MAD* of 5.52% and *RMSE* of 6.65%. Table 1 implies that using the sample period starting from 1970 is a better choice, considering the effect of the empirical return variation.

Why a better prediction occurs by using the sample period 1970 - 2018 than the longer sample time period? As we have argued in the beginning, empirically, the return variation also affect the prediction. To explain this phenomenon, we calculate several statistical measures for different time periods in Table 7. Since our purpose is to predict r_n, or equivalently, $R_n = 1/r_n$, we focus on the historical data of R_n. [6] The first measure is the empirical variance, $\overline{\sigma}^2$ of the sequence $\{R_t\}$. The second is $\Omega \equiv \frac{1}{T}\sum_{t=1}^{T} R_t^2$, representing the average of R_t^2. Lastly, $\zeta_{max} = \max\{|R_t|\}$ representing the maximum jump size of the sequence $\{R_t\}$. As shown in Panel A, Table 7, each measure in the time period 1970-2018 is smaller than or equal to the corresponding measure in the time period 1933-2018, 1946-2018. For instance, $\Omega = 1.0073$ in the time period 1970-2018 while the same measure is 1.0091 and 1.0081 for the time period 1933 - 2018 and 1946 - 2018, respectively.

We also calculate the correlation coefficient ($\rho_{y,x}$) between the stock return and the interest rate, and the regression coefficient ($\hat{\rho}_{y,x}$) of the stock return on the interest rate in Table 7. Each correlation and the regression $\hat{b}$ is negative, and its absolute value in the time period 1970-2018 is always greater than the corresponding one in the sample period from 1933 or 1946. It implies that the empirical variance of $R_t = \frac{y_t}{x_t}$ is the smallest one with the sample period from 1970. Therefore, the prediction is jointly affected by the number of the historical data and the variation of the historical data as shown in Table 1. Our empirical results are consistent with theoretical prediction mentioned at the beginning of this section.

Next, we keep the sample period to 1970-2018 and use the daily data instead. The empirical

[5]We also test whether there is any statistical difference between $\eta = 1$ and other values, a simple t-test fails to reject the null hypothesis.

[6]Notice that the variance of $\frac{y}{x}$ depends on the variance of y and x and the correlation between them. Therefore, the variation of the relative stock return relies on the empirical covariance structure of the interest rate and the stock return.

Table 1: **Prediction of monthly relative return using S&P 500 index**

Reported are various predictive statistics during the period of 2010 - 2018 using monthly relative return of S&P 500 index to 3-month T-Bill. The predictive error is defined as,

$$\text{predictive error} = \frac{\hat{R}_n - R_n^M}{R_n^M},$$

where $\hat{R}_n$ is the predicted relative return using one of predicting formulas and R_n^M is the market observed relative return at time n. Each column corresponds to a particular predicting formula. Panel A uses the full sample to start the prediction; Panel B starts the prediction from 1946, the post WWII period; Panel C starts the prediction from 1970; and Panel D starts the prediction from 2000. In each panel, reported statistics are *MAD*, the average of the absolute predictive errors; *Bias*: the average of the predictive errors; *Deviation*: the average of the squared difference between the bias and the predictive errors; *RMSE*: the square root of the average of the sum of squared predictive errors.

			sextic formula				
	benchmark	**cubic formula**	$\eta = 1$ **(Cover)**	$\eta = 0.75$	$\eta = 0.5$	$\eta = 0.15$	$\eta = 0.05$
Panel A: Data Sample from 1934 to 2018 (N = 1020)							
MAD	54.17%	531.90%	7.36%	7.70%	8.09%	8.73%	8.92%
Bias	-54.17%	94.10%	7.22%	7.58%	7.99%	8.64%	8.84%
Deviation	0.35%	7880.80%	0.17%	0.17%	0.17%	0.18%	0.18%
RMSE	54.49%	892.71%	8.30%	8.63%	9.01%	9.62%	9.81%
Panel B: Data Sample from 1946 to 2018 (N = 876)							
MAD	49.25%	582.48%	9.42%	9.76%	10.15%	10.74%	10.92%
Bias	-49.25%	-179.46%	9.32%	9.68%	10.07%	10.67%	10.85%
Deviation	0.52%	11699.61%	0.20%	0.21%	0.21%	0.22%	0.23%
RMSE	49.78%	1096.43%	10.35%	10.70%	11.08%	11.67%	11.85%
Panel C: Data Sample from 1970 to 2018 (N = 588)							
MAD	22.70%	35.62%	**5.52%**	5.64%	5.77%	5.96%	6.01%
Bias	-8.30%	34.63%	4.83%	4.97%	5.11%	5.32%	5.38%
Deviation	5.62%	10.02%	0.21%	0.21%	0.22%	0.22%	0.23%
RMSE	25.12%	46.92%	6.65%	6.78%	6.92%	7.12%	7.18%
Panel D: Data Sample from 2000 to 2018 (N = 228)							
MAD	64.15%	12.71%	9.92%	9.97%	10.02%	10.08%	10.10%
Bias	60.79%	-6.56%	-4.22%	-4.22%	-4.22%	-4.22%	-4.22%
Deviation	58.27%	1.91%	1.35%	1.36%	1.38%	1.39%	1.40%
RMSE	97.59%	15.29%	12.37%	12.42%	12.47%	12.54%	12.56%

Table 2: **Prediction of daily relative return using S&P 500 index**

Reported are various predictive statistics during the period of 2010 - 2018 using daily relative return of S&P 500 index to 1-year T-Bill. The predictive error is defined as,

$$\text{predictive error} = \frac{\hat{R}_n - R_n^M}{R_n^M},$$

where $\hat{R}_n$ is the predicted relative return using one of predicting formulas and R_n^M is the market observed relative return at time n. Each column corresponds to a particular predicting formula. Panel A uses the full sample starting from 1970 to apply the prediction; Panel B uses the subsample starting from 2000. In each panel, reported statistics are *MAD*, the average of the absolute predictive errors; *Bias*: the average of the predictive errors; *Deviation*: the average of the squared difference between the bias and the predictive errors; *RMSE*: the square root of the average of the sum of squared predictive errors.

			sextic formula				
	benchmark	**cubic formula**	η **= 1 (Cover)**	η **= 0.75**	η **= 0.5**	η **= 0.15**	η **= 0.05**
Panel A: Data Sample from 1970 to 2018 (N = 12228)							
MAD	44.16%	7.19%	**1.96%**	2.00%	2.06%	2.13%	2.15%
Bias	39.00%	4.02%	1.78%	1.83%	1.88%	1.96%	1.98%
Deviation	27.43%	0.56%	0.02%	0.02%	0.02%	0.02%	0.02%
RMSE	65.30%	8.49%	2.29%	2.34%	2.40%	2.48%	2.51%
Panel B: Data Sample from 2000 to 2018 (N = 4744)							
MAD	135.08%	7.77%	2.95%	2.98%	3.01%	3.05%	3.06%
Bias	135.07%	-6.80%	-1.65%	-1.67%	-1.68%	-1.70%	-1.71%
Deviation	262.14%	0.42%	0.12%	0.12%	0.13%	0.13%	0.13%
RMSE	210.62%	9.42%	3.86%	3.90%	3.93%	3.99%	4.00%

results using daily data are remarkable. Table 2 reports various predictive statistics during the period of 2010 - 2018 using daily relative return of S&P 500 index to 1-year T-Bill. As the data frequency increases, both the cubic and sextic prediction formula outperform the benchmark based on standard equilibrium. In Panel A, we use the sample starting from 1970 to start the prediction with a total of 12,228 daily observations. The *MAD* and *RMSE* are about 7.19% and 8.49%, respectively using cubic formula. Excitingly, the *MAD* and *RMSE* using the sextic polynomial formula is as low as 1.96% and 2.29%. In Panel B, we start the prediction using the subsample of 2000 - 2018. With less than half of the data in Panel A, we obtain a slightly larger *MAD* of 2.95% and *RMSE* of 3.86% using the sextic formula with $\eta = 1$.

4.3 Prediction in CRSP and Wilshire 5000 Index

By using CRSP value weighed relative return data, Table 3 and Table 4 report the *MAD*, *Bias*, *Deviation*, and *RMSE* of the predicting monthly and daily prediction, respectively. To show the effect of the return variation to the prediction, Table 7, Panel B reports the statistical measures of the sequence $\{R_t\}$ in several different time periods.

Overall, these results demonstrate the same pattern as we have observed using S&P 500 index. The sextic polynomial formula dominates other methods and the information-theory based formulas (both cubic and sextic) perform better than the benchmark when using daily data. For instance, Panel C of Table 3 document that the *MAD* and *RMSE* using sextic formula are estimated around 5% and 6% for monthly data. Moreover, by using the data from the time period 1970 to 2018, we obtain the best prediction error.

Remarkably, the sextic polynomial formula provide much more accurate forecasts using the daily data. Panel A of Table 4 demonstrate that the mean of absolute forecast error, *MAD*, is bounded strictly by 2%. The lowest *MAD* is about 1.14% when $\eta = 1$. Likewise, the *RMSE* is also bounded strictly by 2%. The lowest *RMSE* is about 1.42% when $\eta = 1$. As we shorten the data sample, the *MAD* and *RMSE*, shown in Panel B, increases to 2.70% and 3.59%. Still, the forecast performance of the sextic polynomial formula is reasonably good.

Finally, we report the prediction performance using monthly and daily Wilshire 5000 index. The statistics are summarized in Table 5 and Table 6 respectively. The results still support the main findings using S&P 500 index and CRSP value weighted index. However, compared with other indices, the errors for Wilshire 5000 become larger even when we use the sextic polynomial formula. Further, the sample period of 1970 - 2018 is no longer the best choice. The errors become

Table 3: **Prediction of monthly relative return using CRSP value weighted index**

Reported are various predictive statistics during the period of 2010 - 2018 using monthly relative return of CRSP value weighted index to 3-month T-Bill. The predictive error is defined as,

$$\text{predictive error} = \frac{\hat{R}_n - R_n^M}{R_n^M},$$

where $\hat{R}_n$ is the predicted relative return using one of predicting formulas and R_n^M is the market observed relative return at time n. Each column corresponds to a particular predicting formula. Panel A uses the full sample to start the prediction; Panel B starts the prediction from 1946, the post WWII period; Panel C starts the prediction from 1970; and Panel D starts the prediction from 2000. In each panel, reported statistics are *MAD*, the average of the absolute predictive errors; *Bias*: the average of the predictive errors; *Deviation*: the average of the squared difference between the bias and the predictive errors; *RMSE*: the square root of the average of the sum of squared predictive errors.

			sextic formula				
	benchmark	**cubic formula**	$\eta = 1$ **(Cover)**	$\eta = 0.75$	$\eta = 0.5$	$\eta = 0.15$	$\eta = 0.05$
			Panel A: Data Sample from 1934 to 2018 (N = 1020)				
MAD	55.54%	800.46%	7.63%	7.98%	8.39%	9.05%	9.26%
Bias	-55.54%	-409.71%	7.50%	7.87%	8.29%	8.97%	9.18%
Deviation	0.23%	31730.00%	0.17%	0.17%	0.18%	0.18%	0.18%
RMSE	55.74%	1827.80%	8.57%	8.91%	9.30%	9.93%	10.13%
			Panel B: Data Sample from 1946 to 2018 (N = 876)				
MAD	47.07%	2124.32%	8.40%	8.71%	9.05%	9.57%	9.73%
Bias	-47.07%	23.96%	8.27%	8.59%	8.94%	9.48%	9.64%
Deviation	0.45%	575893.53%	0.19%	0.19%	0.20%	0.20%	0.21%
RMSE	47.54%	7588.80%	9.34%	9.65%	9.98%	10.49%	10.65%
			Panel C: Data Sample from 1970 to 2018 (N = 588)				
MAD	20.00%	28.77%	**5.06%**	5.17%	5.28%	5.45%	5.50%
Bias	-4.86%	27.85%	4.31%	4.43%	4.57%	4.76%	4.82%
Deviation	4.78%	5.34%	0.19%	0.19%	0.20%	0.20%	0.20%
RMSE	22.40%	36.19%	6.13%	6.24%	6.36%	6.54%	6.59%
			Panel D: Data Sample from 2000 to 2018 (N = 228)				
MAD	43.14%	11.34%	7.77%	7.82%	7.86%	7.93%	7.94%
Bias	39.84%	-4.20%	-1.58%	-1.57%	-1.56%	-1.54%	-1.54%
Deviation	23.77%	1.71%	0.92%	0.92%	0.93%	0.95%	0.95%
RMSE	62.96%	13.75%	9.70%	9.74%	9.79%	9.86%	9.88%

Table 4: **Prediction of daily relative return using CRSP value weighted index**

Reported are various predictive statistics during the period of 2010 - 2018 using daily relative return of CRSP value weighted index to 1-year T-Bill. The predictive error is defined as,

$$\text{predictive error} = \frac{\hat{R}_n - R_n^M}{R_n^M},$$

where $\hat{R}_n$ is the predicted relative return using one of predicting formulas and R_n^M is the market observed relative return at time n. Each column corresponds to a particular predicting formula. Panel A uses the full sample starting from 1970 to apply the prediction; Panel B uses the subsample starting from 2000. In each panel, reported statistics are *MAD*, the average of the absolute predictive errors; *Bias*: the average of the predictive errors; *Deviation*: the average of the squared difference between the bias and the predictive errors; *RMSE*: the square root of the average of the sum of squared predictive errors.

			sextic formula				
	benchmark	**cubic formula**	η **= 1 (Cover)**	η **= 0.75**	η **= 0.5**	η **= 0.15**	η **= 0.05**
Panel A: Data Sample from 1970 to 2018 (N = 12228)							
MAD	180.80%	5.82%	**1.14%**	1.16%	1.18%	1.20%	1.21%
Bias	180.80%	-5.21%	0.23%	0.24%	0.25%	0.26%	0.26%
Deviation	345.16%	0.22%	0.02%	0.02%	0.02%	0.02%	0.02%
RMSE	259.24%	7.03%	1.42%	1.44%	1.46%	1.49%	1.50%
Panel B: Data Sample from 2000 to 2018 (N = 4744)							
MAD	118.28%	7.89%	2.70%	2.73%	2.75%	2.79%	2.80%
Bias	118.28%	-7.62%	-1.91%	-1.93%	-1.95%	-1.98%	-1.98%
Deviation	114.91%	0.29%	0.09%	0.09%	0.10%	0.10%	0.10%
RMSE	159.62%	9.31%	3.59%	3.62%	3.65%	3.70%	3.71%

smaller when we use the data sample from 2000 to 2018. This finding is not a surprise. As we have mentioned before, the index level of Wilshire 5000 index is very small, and the earlier data quality is poor. The data quality issue may hinder the performance of the predicting formula.

Despite the uncertainty from monthly data, the sextic polynomial predicting formula is still promising when we use the daily data of Wilshire 5000 index. As noted in Table 5, the *MAD* and *RMSE* using the full sample of 1981 - 2018 are about 12.96% and 13.27%. Importantly, the *MAD* and *RMSE* in Table 6 are dramatically reduced to 3.71% and 4.33% when using the subsample of 2000 - 2018. This evidence is consistent with our earlier argument that the predicting performance depends on both the number of historical data as well as the empirical covariance structure.

5 Conclusion

In this paper, we propose a data-driving prediction methodology for the relative stock return given sufficient historical data. This methodology depends on several on-line strategies in universal portfolio literature. Our empirical testing is supportive, and the relative prediction error depends on the empirical covariance structure and the quality of historical market data. This prediction methodology can be extended to multiple assets prediction by using several universal portfolios simultaneously.

Table 5: **Prediction of monthly relative return using Wilshire 5000 index**

Reported are various predictive statistics during the period of 2010 - 2018 using monthly relative return of Wilshire 5000 index to 3-month T-Bill. The predictive error is defined as,

$$\text{predictive error} = \frac{\hat{R}_n - R_n^M}{R_n^M},$$

where $\hat{R}_n$ is the predicted relative return using one of predicting formulas and R_n^M is the market observed relative return at time n. Each column corresponds to a particular predicting formula. Panel A uses the full sample starting from 1971 to apply the prediction; Panel B uses the subsample starting from 2000. In each panel, reported statistics are *MAD*, the average of the absolute predictive errors; *Bias*: the average of the predictive errors; *Deviation*: the average of the squared difference between the bias and the predictive errors; *RMSE*: the square root of the average of the sum of squared predictive errors.

			sextic formula				
	benchmark	**cubic formula**	η **= 1 (Cover)**	η **= 0.75**	η **= 0.5**	η **= 0.15**	η **= 0.05**
Panel A: Data Sample from 1971 to 2018 (N = 575)							
MAD	60.17%	134.76%	28.28%	29.91%	31.35%	33.66%	33.65%
Bias	-60.17%	-37.66%	28.28%	29.91%	31.35%	33.66%	33.65%
Deviation	0.28%	1271.30%	1.21%	1.44%	1.54%	2.01%	1.83%
RMSE	60.40%	358.54%	30.34%	32.24%	33.71%	36.52%	36.27%
Panel B: Data Sample from 2000 to 2018 (N = 228)							
MAD	29.59%	44.50%	18.12%	18.43%	18.77%	19.28%	19.43%
Bias	2.24%	36.38%	13.44%	13.74%	14.06%	14.55%	14.70%
Deviation	12.56%	43.33%	4.06%	4.25%	4.46%	4.78%	4.89%
RMSE	35.51%	75.21%	24.23%	24.77%	25.36%	26.27%	26.55%

Table 6: **Prediction of daily relative return using Wilshire 5000 index**

Reported are various predictive statistics during the period of 2010 - 2018 using daily relative return of Wilshire 5000 index to 1-year T-Bill. The predictive error is defined as,

$$\text{predictive error} = \frac{\hat{R}_n - R_n^M}{R_n^M},$$

where $\hat{R}_n$ is the predicted relative return using one of predicting formulas and R_n^M is the market observed relative return at time n. Each column corresponds to a particular predicting formula. Panel A uses the full sample starting from 1981 to apply the prediction; Panel B uses the subsample starting from 2000. In each panel, reported statistics are *MAD*, the average of the absolute predictive errors; *Bias*: the average of the predictive errors; *Deviation*: the average of the squared difference between the bias and the predictive errors; *RMSE*: the square root of the average of the sum of squared predictive errors.

			sextic formula				
	benchmark	**cubic formula**	η **= 1 (Cover)**	η **= 0.75**	η **= 0.5**	η **= 0.15**	η **= 0.05**
Panel A: Data Sample from 1981 to 2018 (N = 9379)							
MAD	54.74%	132.56%	12.96%	13.32%	13.72%	14.34%	14.53%
Bias	-54.74%	132.56%	12.96%	13.32%	13.72%	14.34%	14.53%
Deviation	0.40%	48.92%	0.08%	0.09%	0.10%	0.11%	0.11%
RMSE	55.11%	149.88%	13.27%	13.65%	14.07%	14.72%	14.92%
Panel B: Data Sample from 2000 to 2018 (N = 4737)							
MAD	38.65%	9.72%	**3.71%**	3.76%	3.81%	3.89%	3.91%
Bias	24.40%	3.79%	2.23%	2.27%	2.31%	2.36%	2.38%
Deviation	25.10%	1.16%	0.14%	0.14%	0.15%	0.15%	0.15%
RMSE	55.72%	11.42%	4.33%	4.40%	4.46%	4.55%	4.57%

Table 7: **Empirical volatility structure of relative return**

This table reports various statistics to measure the variation of the relative return structure during different time periods. Three measures, σ^2, Ω, and ζ_{max} are defined as follows,

$$\sigma^2 = \frac{1}{N-1}\sum_{n=1}^{N}(R_n - \bar{R})^2, \quad \Omega = \frac{1}{N}\sum_{n=1}^{N} R_n^2, \quad \text{and} \quad \zeta_{max} = \max(|R_n|),$$

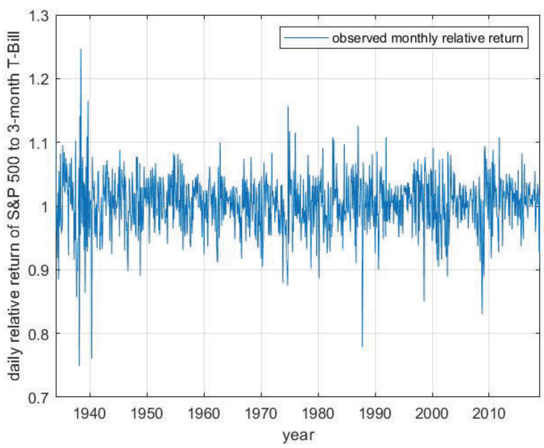
where R_n is calculated as $R_n = \frac{y_n}{x_n}$. In the last two columns, we report the regression coefficient $\hat{\beta}_{y,x}$ of y_n on x_n, and the correlation coefficient, $\rho_{y,x}$, between these two. Panel A uses monthly return of S&P 500 index to proxy y_n while Panel B uses the monthly CRSP value weighted index. For both panels, x_n is calculated as one plus 3-month T-Bill rate.

Panel A: S&P 500 index

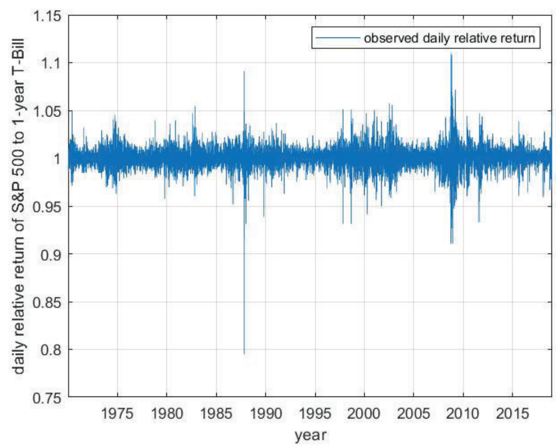
Period	$\bar{\sigma}^2$	Ω	ζ_{max}	$\hat{\beta}_{y,x}$	$\rho_{y,x}$
1933-2018	0.0020	1.0091	1.2470	-0.4512	-0.0264
1946-2018	0.0017	1.0081	1.1559	-0.6121	-0.0383
1970-2018	0.0017	1.0073	1.1559	-0.6307	-0.0396
2000-2018	0.0018	1.0064	1.1559	-0.5275	-0.0336

Panel B: CRSP value weighted index

Period	$\bar{\sigma}^2$	Ω	ζ_{max}	$\hat{\beta}_{y,x}$	$\rho_{y,x}$
1933-2018	0.0020	1.0092	1.2324	-0.4789	-0.0279
1946-2018	0.0019	1.0080	1.1553	-0.5961	-0.0365
1970-2018	0.0019	1.0075	1.1553	-0.6689	-0.0407
2000-2018	0.0019	1.0067	1.1553	-0.5686	0.0350



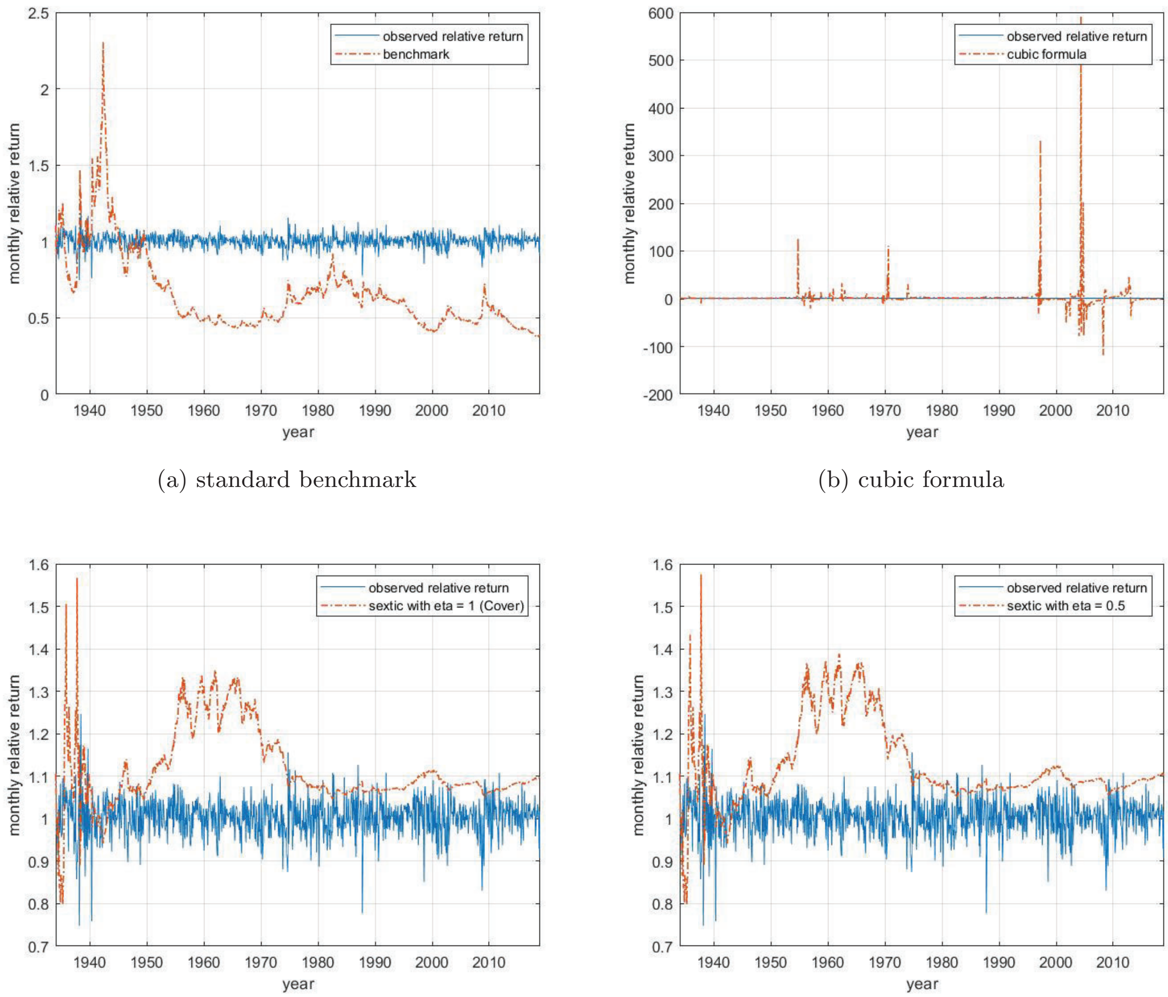
(a) monthly relative return



(b) daily relative return.

Figure 1: **Relative return of S&P 500 index to T-Bill**

Figure 1 plot the monthly and daily relative return of S&P 500 index to T-Bill with respect to time periods using the formula $R_n = \frac{y_n}{x_n}$. y_n is the return of S&P 500 index. x_n is the return of 3-month T-bill for monthly data (Figure 1(a)) and of 1-year T-bill for daily data (Figure 1(b)).



(a) standard benchmark

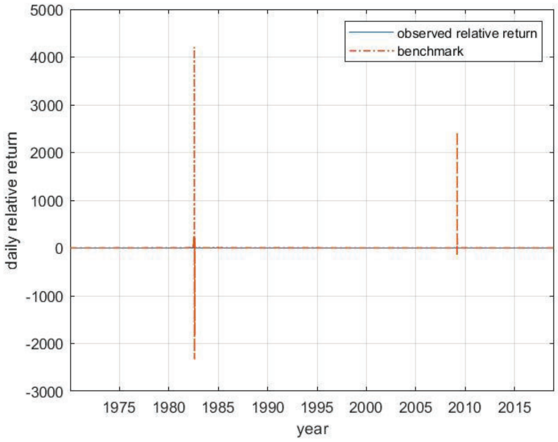
(b) cubic formula

(c) sextic formula with $\eta = 1$

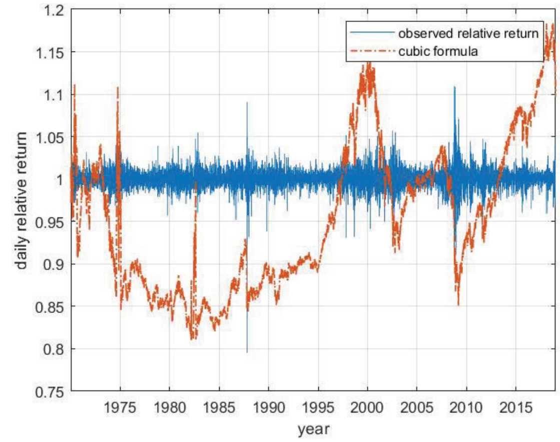
(d) sextic formula with $\eta = 0.5$

Figure 2: **Prediction of monthly relative return using S&P 500 index**

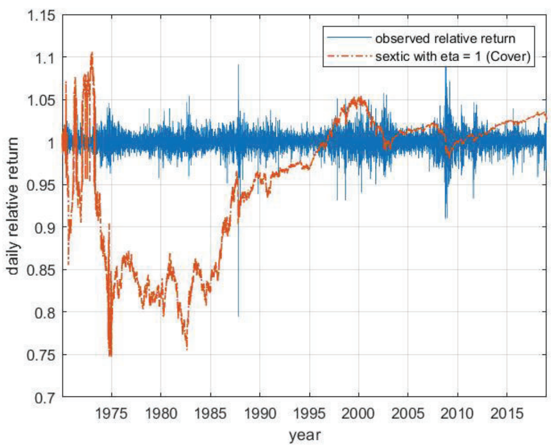
Figure 2 compares the three predicting formulas with respect to monthly market observed relative return of S&P 500 index to 3-month T-Bill. Figure 2(a) plots the predicted return using the benchmark formula in Equation (2); Figure 2(b) plots the predicted return using the cubic formula in Equation (12); Figure 2(c) plots the predicted return using the sextic formula by Cover in Equation (12); and Figure 2(d) plots the predicted return using the sextic formula in Equation (2) with an alternative $\hat{b}$ construction of Equation (13) where $\eta = 0.5$.



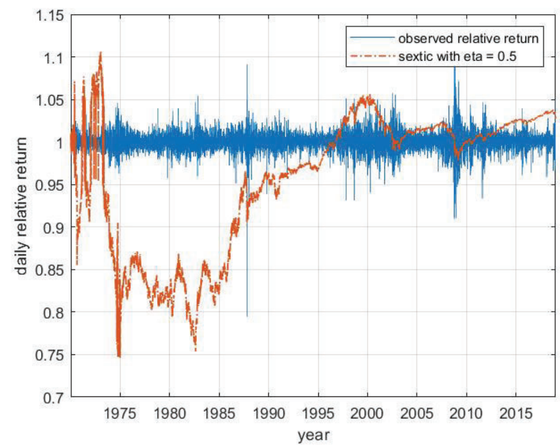
(a) standard benchmark



(b) cubic formula



(c) sextic formula with $\eta = 1$



(d) sextic formula with $\eta = 0.5$

Figure 3: **Prediction of daily relative return using S&P 500 index**

Figure 3 compares the three predicting formulas with respect to monthly market observed relative gross return of S&P 500 index to 3-month T-Bill. Figure 3(a) plots the predicted return using the benchmark formula in Equation (2); Figure 3(b) plots the predicted return using the cubic formula in Equation (12); Figure 3(c) plots the predicted return using the sextic formula by Cover in Equation (12); and Figure 3(d) plots the predicted return using the sextic formula in Equation (2) with an alternative $\hat{b}$ construction of Equation (13) where $\eta = 0.5$.

References

Algoget, P. (1992): Universal Schemes for Prediction, Gambling and Portfolio Selecton, *Annals of Probability*, 20 (2), 901-941.

Breimain, L. (1961): Optimal Gambling Systems for Favorable Games, *Proceeding Fourth Berkelep Sumpo. Math. Statist. Probab* 1, 65-78.

Cesa-Bianchi, N., and G. Lugosi. (2006): *Prediction, Learning, and Games*, Cambridge University Press.

Cremers, K. M. (2002): Stock Return Predictability: A Bayesian Model Selection Perspective, *Review of Financial Studies* 15(4), 1223-1249.

Cochrane, J. H. (2000): *Asset Pricing*, Princeton University Press.

Cover, T. (1991): Universal Portfolios, *Mathematical Finance* 1, 1-29.

Cover, T., and J. Thomas (1991): *Elements of Information Theory*, Wiley.

Foster, D., and R. Vohra (1999): Regret in the On-Line Decision Problem, *Games and Economic Behavior* 29, 7-35.

Hastie, T., Tibshirani, R., and J. Friedman (2016): *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, Second Edition, Springer Series in Statistics, Springer.

Hazan, E., A. Agarwal, and S. Kale. (2007): Logarithmic Regret Algorithms for Onine Convex Optimization, *Macheine Learning* 69 (2-3), 169-192.

Hazan, E, and S. Kale (2015): An Online Portfolio Selection Algorithm with Regret Logarithmic in Price Variation, *Mathematical Finance* 25 (2), 288-310.

Helmbold, D. P., R. E. Schapire, Y. Singer, and M. K. Warmuth. (1998): On-line Portfolio Selection Using multiplicative Updates, *Mathemtical Finance* 8 (4), 325-347.

Vovk, V, abd C. Watkins. (1998): Universal Portfolio Selection, *Proceedings of the Eleventh Annual Conference in Computational Learning Theory*, New York; ACM press, 12-23.

Appendix A

To prove Proposition 2.1, we start with two simple lemmas.

Lemma 5.1. *For any market data, the best constant proportion strategy b_N^* and all $N \geq 1$,*

$$\frac{\partial b_N^*}{\partial r_N} > 0. \tag{A-1}$$

Proof. We compute the partial derivative of both sides in equation (7) to r_N, obtaining

$$\sum_{n=1}^{N-1} -\frac{r_n^2}{(1+r_n b_N^*)^2}\frac{\partial b_N^*}{\partial r_N} + \frac{1 - r_N^2 \frac{\partial b_N^*}{\partial r_N}}{(1+r_N b_N^*)^2} = 0.$$

Reorganizing this last equation, we obtain

$$\sum_{n=1}^{N} \frac{r_n^2}{(1+r_n b_N^*)^2}\frac{\partial b_N^*}{\partial r_N} = \frac{1}{(1+r_N b_N^*)^2}.$$

Therefore, $\frac{\partial b_N^*}{\partial r_N} > 0$. The proof is finished. □

Lemma 5.2. *Given a sequence of function $\{f_N(x,y); N = 1,2,\cdots,\}$ where x takes value in a compact subset $\mathcal{C}$ in $\mathbb{R}^d$ and y takes values in $[m,M]$. Assume that $f_N(x,g_N(x)) = o(1)$ for all x for a certain function $g_N(x)$, and there exists a positive number c such that*

$$|\frac{\partial f_N(x,y)}{\partial y}| \geq c, \forall (x,y) \in \mathcal{C} \times [m,M], N \geq 1$$

Define $h_N(x) \in [m,M]$ by the following equation

$$f_N(x,h_N(x)) = 0.$$

Then, $g_N(x) - h_N(x) = o(1)$.

Proof. By the mean-value theorem, we have

$$|f_N(x,y) - f_N(x,z)| = |y-z| \left|\frac{\partial f_N(x,y)}{\partial y}\right|_{y=\zeta} \geq c|y-z|$$

where ζ between y and z. Letting $y = g_N(x)$ and $z = h_N(x)$, then

$$c|g_N(x) - h_N(x)| \leq |f_N(x,g_N(x)) - f_N(x,h_N(x))| = |f_N(x,g_N(x))| = o(1).$$

Hence, $g_N(x) - h_N(x)| = o(1)$, The proof is completed. □

Proof of Proposition 2.1.

To apply Lemma 5.2, we define, for $x = (r_1, \cdots, r_{N-1})$ and $y = r_N$,

$$f_N(x,y) \equiv \frac{1}{N}\left(\sum_{n=1}^{N} log(1+b^* r_n) - \sum_{n=1}^{N} log(1+\hat{b}_N r_N)\right) - \frac{log(N+1)}{2N}.$$

By assumption, $(x, y) = (r_1, \cdots, r_N)$ takes values in a compact subset of a Euclid space. By Cover's theorem, r_N can be viewed as $g_N(x)$ since $f_N(x, r_N) = o(1)$. We use $h_N(x)$ to interpret the "prediction" of r_N in solving Equation (8) and Equation (7). Since Lemma 5.2 concerns on one equation only, we plug $b^*_N(x, y)$ into Equation (8). By Lemma 5.1 and the implicit function theorem, $b^*_N(x, y)$ is well defined. Therefore, Lemma 5.2 can be applied in this situation.

It suffices to show that $|\frac{\partial f_N(x,y)}{\partial y}|$ is uniformly bounded from below by a positive number for all $N \geq 1$ and all (x, y). By straightforward calculation, we have (with the notation $b^* = b^*_N$)

$$\begin{aligned} \frac{\partial f_N(x,y)}{\partial y} &= \frac{1}{N}\left(\sum_{n=1}^{N-1} \frac{r_n}{1+b^* r_n}\frac{\partial b^*}{\partial y} + \frac{1}{1+b^* y}\frac{\partial (b^* y)}{\partial y}\right) \\ &\quad -\frac{1}{N}\frac{\hat{b}_N}{1+\hat{b}_N y} \end{aligned}$$

in which we make use of the on-line feature of $\hat{b}$ which depends on available data $r_1, \cdots, r_{N-1}$. Since

$$\frac{\partial (b^* y)}{\partial y} = b^* + \frac{\partial b^*}{\partial y},$$

and $\frac{1}{N}\frac{\hat{b}_N}{1+\hat{b}_N y} \leq \frac{1}{N}$, we have

$$\begin{aligned} \frac{\partial f_N(x,y)}{\partial y} &\geq \frac{1}{N}\sum_{n=1}^{N} \frac{r_n}{1+r_n b^*}\frac{\partial b^*}{\partial y} + \frac{1}{N}\frac{b^*}{1+b^* y} - \frac{1}{N} \\ &\geq \frac{1}{N}\sum_{n=1}^{N} \frac{r_n}{1+r_n b^*}\frac{\partial b^*}{\partial y} - \frac{1}{N} \\ &\geq \min_{1\leq i\leq N}\left\{\frac{r_n}{1+r_n b^*}\right\}\frac{\partial b^*}{\partial y} - \frac{1}{N}. \end{aligned}$$

Since each $\frac{r_n}{1+r_n b^*} \geq \frac{m}{1+mb^*} \geq \frac{m}{1+m}$, and by the assumption that $\frac{\partial b^*}{\partial y} \geq c$, we obtain

$$\frac{\partial f_N(x,y)}{\partial y} \geq \frac{m}{1+m}c - \frac{1}{N}.$$

As $1/N = o(1)$, we see that $|\frac{\partial f_N(x,y)}{\partial y}|$ is clearly uniformly bounded from below for all $N \geq 1$ and all possible data (x, y). The proof is finished. □

Derivation of the cubic prediction formula.

By using the first-order approximation of the function $\log(1+x)$, $\log(1+x) \sim x$, Equation (8) can be simplified into,

$$\sum_{n=1}^{N-1}(b^\star - \hat{b}_n)r_n + (b^\star - \hat{b}_N)r_N - \frac{\log(N+1)}{2} \sim 0. \tag{A-2}$$

Next, by using the first-order Taylor expansion of $\frac{1}{1+x}$, Equation (7) can be simplified into,

$$\sum_{n=1}^{N} r_n(1 - b^\star r_n) \sim 0, \tag{A-3}$$

yielding

$$b_N^\star = \frac{\sum_{n=1}^N r_n}{\sum_{n=1}^N r_n^2}. \tag{A-4}$$

Equivalently,

$$b_N^\star = \frac{P_1 + r_N}{P_2 + r_N^2} \tag{A-5}$$

which depends on historical data as well as r_N.

Plugging the last formula of $b_N^\star$ into Equation (A-2), we obtain,

$$\sum_{n=1}^{N-1}\left(\frac{P_1 + r_N}{P_2 + r_N^2} - \hat{b}_n\right) r_n + \left(\frac{P_1 + r_N}{P_2 + r_N^2} - \hat{b}_N\right) r_N - \frac{\log(N+1)}{2} \sim 0. \tag{A-6}$$

Alternatively,

$$P_1^2 + Q_1 P_2 + (2P_1 - P_2\hat{b}_N)r_N + (1 - Q_1)r_N^2 - \hat{b}_N r_N^3 \sim P_2\frac{\log(N+1)}{2} + r_N^2\frac{\log(N+1)}{2}. \tag{A-7}$$

By reorganizing these terms, we see that r_N solves the cubic function $f_3(x)$ in the approximate sense, $f_3(r_N) \sim 0$. We choose the real root, r_N^{FF}, that is most close to r_{N-1}. The predicted relative stock return from the cubic predicting formula is

$$R_N^{FF} = \frac{1}{1 + r_N^{FF}}. \tag{A-8}$$

□

Derivation of the sextic prediction formula.

By using the second-order Taylor expansion of the function $\log(1+x)$, $\log(1+x) \sim x - \frac{x^2}{2}$, Equation (6) can be simplified into,

$$\sum_{n=1}^{N-1}\left[(b^\star - \hat{b}_n)r_n - \frac{1}{2}(b^{\star 2} - \hat{b}_n^2)r_n^2\right] + (b^\star - \hat{b}_N)r_N - \frac{1}{2}(b^{\star 2} - \hat{b}_N^2)r_N^2 - \frac{\log(N+1)}{2} \sim 0. \tag{A-9}$$

Plugging Equation (A-5), the left side of the last equation becomes we obtain,

$$
\begin{aligned}
&\sum_{n=1}^{N-1} \left\{ \left(\frac{P_1 + r_N}{P_2 + r_N^2} - \hat{b}_n \right) r_n - \frac{1}{2} \left[\left(\frac{P_1 + r_N}{P_2 + r_N^2} \right)^2 - \hat{b}_n^2 \right] r_n^2 \right\} \\
&+ \left(\frac{P_1 + r_N}{P_2 + r_N^2} - \hat{b}_N \right) r_N - \frac{1}{2} \left[\left(\frac{P_1 + r_N}{P_2 + r_N^2} \right)^2 - \hat{b}_N^2 \right] r_N^2 - \frac{\log(N+1)}{2} \sim 0.
\end{aligned}
$$

By straightforward calculation, we derive that $f_6(r_N) \sim 0$. Based on our empirical testing, we can always find numerical real root of the function $f_6(x) = 0$. Formally, it can be formulated as an optimization problem

$$
\min_{r_N : |r_N - r_{N-1}| \leq b} f_6(r_N)^2
$$

where b is one control parameter. The predicted relative stock return from the sextic predicting formula is thus

$$
R_N^{SF} = \frac{1}{1 + r_N^{SF}}. \tag{A-10}
$$

$\square$