The Bregman-divergence universal portfolio associated with a convex polynomial

Choon Peng Tan¹ᵃ and Yap Jia Lee¹ᵇ

¹Department of Mathematical and Actuarial Sciences, LKC Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Jalan Sungai Long, Bandar Sungai Long, 43000 Kajang, Selangor, Malaysia
E-mail: ¹tancp@utar.edu.my, ¹yjlee@utar.edu.my

Abstract. A convex polynomial with a non-integer degree is used to generate a Bregman-divergence universal portfolio together with an application of a first-order binomial expansion of the multiplicative update. The next-day portfolio is a weighted combination of the present-day portfolio and a power of the present-day portfolio, where the weights are depending on functions of the present-day stock price relatives. A parameter delta is introduced in the portfolio with aims to find the parameter that exhibiting the best performance locally. The portfolio is run over five stock data sets selected from the local stock exchange to study and to compare its empirical performance. The empirical results are compared with the results obtained by using the reverse Helmbold universal portfolio. Better performance is demonstrated for the portfolio introduced in this paper with respect to certain values of the parameters used on three data sets. There is promising evidence that this portfolio can increase the investment wealth significantly with the proper choice of the portfolio parameters.

1. Introduction

Markowitz developed the portfolio investment theory in mathematical terms [1] and the Markowitz’s model was then extended by Sharpe on portfolio analysis [2]. Logarithmic utility has been extensively discussed since Daniel Bernoulli’s article on log utility [3]. Kelly discussed the idea of using log utility function in gambling [4]. Latané had, on the other hand, provided an intuitive economic analysis to the finance world using log utility as an investment criterion independent of Kelly’s work [5]. Inspired by Kelly’s gambling point of view, Breiman argued that an optimal gambling system had to consider the minimal time requirement and the magnitude condition [6].

Thorp discussed general optimal betting theory and the theory applied on several favorable games or investments [7]. Thorp had shown that the Kelly strategies were not necessarily mean-variance efficient [8] while Markowitz argued that the Kelly strategy was the limiting mean-variance portfolio [9]. Bell and Cover proved that the expected log-optimum portfolio performed well in either short-term or long-term investment in the game theoretical point of view [10]. An algorithm for maximizing the expected log investment return is presented by Cover in [11]. Cover and Gluss presented an investment scheme with universal properties by using the Bayes decision rules and Blackwell’s approach-exclusion theorem in [12].

Barron and Cover showed that the side information of the underlying stock market distribution is always bound to increment in exponential growth of wealth [13]. Algoet and Cover demonstrated that with probability one for any ergodic market, the conditional expected log return maximization given...
available historical information is asymptotically optimal with no restrictions on the distribution [14].

Extended from the Algoet and Cover asymptotic optimality principle introduced in [14], Algoet discussed universal schemes for prediction, gambling and portfolio selection in the set of all stationary ergodic market with unknown distribution [15]. Cover proved that the wealth achieved by his universal portfolio algorithm asymptotically outperformed the best constant rebalanced portfolio [16]. His analysis was greatly refined in Cover and Ordentlich of which they introduced the idea of side information and generalized Cover’s algorithm to use the uniform (Dirichlet (1,…,1)) and Dirichlet (1/2, … , 1/2) universal portfolios [17].

With an attempt to reduce tremendous consumption in computation time and memory amount for the application of Cover-Ordentlich universal portfolio, the Helmbold universal portfolio was introduced in [18]. An investment portfolio is said to be universal if no stochastic model is assumed for the stock prices. The use of two-stock universal portfolios for investing in the New York Stock Exchange is described in [18], indicating that the universal portfolio is more useful in creating new wealth in long-term investment. Applications of the Kullback-Leibler divergence in generating a universal portfolio is introduced in [18]. A study in [19] showed that the performance of the Helmbold universal portfolio is well-matched to the constant rebalanced portfolio by using some small values of the parameter. The performance of Helmbold universal portfolio can be further improved by using large values of parameters, including some negative values, is demonstrated in [20].

The divergence of two probability distributions is studied widely in the literature, especially in information theory and statistical inference [21]. Two classes of well-known divergences are the $f$-divergence and the Bregman divergence. This study led to the use of the distance function in generating universal portfolios. The reverse Kullback-Leibler divergence generated universal portfolio is studied in [22]. The generalization of Helmbold universal portfolio by scaling the Kullback-Leibler divergence generating the portfolio is studied in [23]. Some connection between the Renyi and the Kullback-Leibler divergence generated universal portfolios is presented in [23]. A study of the general Bregman divergence universal portfolio is done in [24]. The focus in this paper is on a special Bregman-divergence universal portfolio generated by a convex polynomial.

Some basic definitions are in order prior to the presentation of the main result. Investment in a market of $m$ stocks is considered where no stochastic model of the stock prices is assumed. A portfolio vector $\mathbf{b}_n = (b_n)$ used on the $n^{th}$ trading day is a collection of the proportions of the investor’s wealth distributed over the $m$ stocks, namely, $b_n$ is the proportion of the current wealth $S_n$ invested on the $i^{th}$ stock for $i = 1, 2, \ldots, m$, where $0 \leq b_n \leq 1$ and $\sum_{j=1}^{m} b_{nj} = 1$, $n = 1, 2, \ldots$. The market is described by the price-relative vector $\mathbf{x}_n = (x_n)$ on the $n^{th}$ trading day, where $x_{ni}$ is the price-relative of the $i^{th}$ stock on the $n^{th}$ trading day which is defined as the ratio of the closing price of the $i^{th}$ stock to its opening price, for $i = 1, 2, \ldots, m$. The wealth at the end of the $n^{th}$ trading day, $S_n$ is calculated as:

$$S_n = \prod_{j=1}^{n} b_j^i x_j, \ n = 1, 2, \ldots,$$

where $b_j^i x_j = \sum_{j=1}^{m} b_{nj} x_{nj}$ and the initial wealth is assumed to be 1 unit.

Let $f(t)$ be a continuously differentiable convex function. The Bregman divergence of two probability vectors $\mathbf{p} = (p_j)$ and $\mathbf{q} = (q_j)$ generated by $f(t)$ is:

$$B'(\mathbf{p} \parallel \mathbf{q}) = \sum_{j=1}^{m} \left[ f(p_j) - f(q_j) - f'(q_j)(p_j - q_j) \right],$$

(see [21]). Thus, the Bregman divergence of two portfolio vectors $\mathbf{b}_{n+1}$ and $\mathbf{b}_n$ is given by:

$$B'(\mathbf{b}_{n+1} \parallel \mathbf{b}_n) = \sum_{j=1}^{m} \left[ f(b_{n+1,j}) - f(b_{nj}) - f'(b_{nj})(b_{n+1,j} - b_{nj}) \right].$$

(3)
The function \( f(t) = \pm \left[ t^\alpha - at \right] \) for \( t > 0 \) has derivatives

\[
 f'(t) = \pm \alpha t^{\alpha-1} - \alpha \]

and \( f''(t) = \pm \alpha(\alpha-1) t^{\alpha-2} \). It is clear that \( f(t) = t^\alpha - at \), \( t > 0 \) is convex for \( \alpha > 1 \) or \( \alpha < 0 \), whereas \( f(t) = -t^\alpha + at \), \( t > 0 \) is convex for \( 0 < \alpha < 1 \).

2. Main Results

The next-day portfolio \( b_{n+1} \) is to be determined given the current portfolio \( b_n \) and current price-relative \( x_n \). The method is to maximize a certain objective function of \( b_{n+1} \) and determine a stationary point.

The rate of daily wealth increase \( \log(b'_{n+1} x_{n+1}) \) given \( x_n \) and \( b_n \) is to be estimated. Since \( b_{n+1} \) is unknown, \( b_{n+1} \) is estimated roughly as \( x_n \). The first-order Taylor approximation of \( \log(b'_n x_n) \) is

\[
 \log(b'_n x_n) + \frac{b'_n x_n}{b'_n x_n} - 1 .
\]

The distance or divergence \( B'(b_{n+1} || b_n) \) is to be minimized so that \( b_{n+1} \) is close to \( b_n \). Thus, a certain objective function \( \hat{F}(b_{n+1}; \lambda) \) is to be maximized, namely \( \hat{F}(b_{n+1}; \lambda) \), as a linear combination of \( \log(b'_n x_n) + \frac{b'_n x_n}{b'_n x_n} - 1 \) and \(-B'(b_{n+1} || b_n)\) subject to the constraint

\[
 \sum_{j=1}^m b_{n+1,j} = 1.
\]

Proposition 1. Consider the objective function

\[
 \hat{F}(b_{n+1}; \lambda) = \xi \left[ \log(b'_n x_n) + \frac{b'_n x_n}{b'_n x_n} - 1 \right] - B'(b_{n+1} || b_n) + \lambda \left[ \sum_{j=1}^m b_{n+1,j} - 1 \right],
\]

where \( \xi > 0 \) and \( \lambda \) is the Lagrange multiplier. The universal portfolio generated by the convex function \( f(t) = \pm \left[ t^\alpha - at \right] \), \( t > 0 \) is given by:

\[
 b_{n+1,i} = \left[ b_{n,i}^{\alpha-1} + \left( \frac{\xi}{\pm \alpha} \left( \frac{x_n}{b'_n x_n} \right) + \left( \frac{\eta - \xi}{\pm \alpha} \right) \right]^{\frac{1}{\alpha-1}} \text{ for } i = 1, 2, \ldots, m,
\]

where \( \eta \) is another parameter.

Proof. Consider \( f(t) = t^\alpha - at \), \( t > 0 \). Substituting the value of \( f(t) \) in (3) leads to

\[
 B'(b_{n+1} || b_n) = \sum_{j=1}^m \left[ b_{n+1,j}^{\alpha-1} - ab_{n+1,j} - b_{n,j} + ab_{n,j} - ab_{n,j}^{-1} b_{n+1,j} + ab_{n,j} + ab_{n+1,j} - ab_{n,j} \right].
\]

Then,

\[
 B'(b_{n+1} || b_n) = \sum_{j=1}^m \left[ b_{n+1,j}^{\alpha-1} - b_{n,j} - ab_{n,j}^{-1} b_{n+1,j} + ab_{n,j} + ab_{n+1,j} - ab_{n,j} \right]
 = \sum_{j=1}^m \left[ b_{n+1,j}^{\alpha-1} - b_{n,j} - ab_{n,j}^{-1} b_{n+1,j} + ab_{n,j} \right]
 = \sum_{j=1}^m \left[ b_{n+1,j}^{\alpha-1} + (\alpha - 1) b_{n,j} - ab_{n,j}^{-1} b_{n+1,j} \right]
 = \sum_{j=1}^m \left[ b_{n+1,j}^{\alpha-1} - \left( \alpha - 1 \right) b_{n+1,j} \right].
\]

Hence,

\[
 B'(b_{n+1} || b_n) = \sum_{j=1}^m \left[ b_{n+1,j}^{\alpha-1} + (\alpha - 1) b_{n,j} - ab_{n+1,j} \right].
\]
For the objective function (5) where \( B'(b_{n+1} \| b_n) \) is defined by (7), the derivative is given by:

\[
\frac{\partial \hat{F}}{\partial b_{n+1,i}} = \xi \left[ \frac{x_{ni}}{b'_n x_n} \right] - \alpha \left[ b_{n+1,i}^{a-1} - b_{it}^{a-1} \right] + \lambda = 0 \quad \text{for } i = 1, 2, \ldots, m .
\]  

(8)

Multiply (8) by \( b_{ni} \) and sum over \( i \) to obtain

\[
\xi - \alpha \sum_{j=1}^{m} b_{nj} b_{n+1,j}^{a-1} - \sum_{j=1}^{m} b_{nj} = \lambda = 0 .
\]  

(9)

Define

\[
\eta = \alpha \left[ \sum_{j=1}^{m} b_{nj} b_{n+1,j}^{a-1} - \sum_{j=1}^{m} b_{nj} \right]
\]  

(10)

and from (9)

\[
\lambda = -\xi + \eta .
\]  

(11)

Substitute the value of \( \lambda \) in (11) into (8) to obtain

\[
b_{n+1,i} = \left[ b_{ni}^{a-1} + \left( \frac{\xi}{\alpha} \right) \left( \frac{x_{ni}}{b'_n x_n} \right) + \left( \frac{\eta - \xi}{\alpha} \right) \right]^{\frac{1}{a-1}}\text{ for } i = 1, 2, \ldots, m .
\]  

(13)

For \( f(t) = -\left[ t^a - \alpha t \right] \), \( t > 0 \), (12) is again obtained where \( \alpha \) is replace by \( -\alpha \).

Proposition 2. The pseudo Bregman-divergence universal portfolio generated by the convex function \( f(t) = \pm\left[ t^a - \alpha t \right] \), \( t > 0 \) can be chosen as:

\[
b_{n+1,i} = \frac{b_{ni}^{a-1} + \delta b_{ni}^{a-\alpha} \left( \frac{x_{ni}}{b'_n x_n} \right)}{\sum_{j=1}^{m} b_{nj}^{a-1} + \delta b_{nj}^{a-\alpha} \left( \frac{x_{nj}}{b'_n x_n} \right)} \quad \text{for } i = 1, 2, \ldots, m .
\]  

(13)

where \( \delta > 0 \) and \( \alpha \) is real such that \( \alpha \neq 0 \) and \( \alpha \neq 1 \).

Proof. First note that, since \( f^*(t) = \pm\left[ \alpha (\alpha - 1) t^{a-2} \right] > 0 \), \( t > 0 \), this implies that \( \pm\left[ \alpha (\alpha - 1) \right] > 0 \) for all \( \alpha \neq 0,1 \). Thus \( \alpha \neq 0 \). From (6), choosing \( \eta = \xi \) leads to

\[
b_{n+1,i} = \left[ b_{ni}^{a-1} + \left( \frac{\xi}{\pm \alpha} \right) \left( \frac{x_{ni}}{b'_n x_n} \right) \right]^{\frac{1}{a-1}} = b_{ni} \left[ 1 + \left( \frac{\xi}{\pm \alpha} \right) \left( \frac{x_{ni}}{b_{ni}^{a-1} \left( b'_n x_n \right)} \right) \right]^{\frac{1}{a-1}} .
\]  

(14)

By a first-order binomial approximation of the power on the right-hand side,

\[
b_{n+1,i} = b_{ni} \left[ 1 + \left( \frac{\xi}{\pm \alpha (\alpha - 1)} \right) \left( \frac{x_{ni}}{b_{ni}^{a-1} \left( b'_n x_n \right)} \right) \right] .
\]  

(15)

Letting \( \delta = \pm \frac{\xi}{\alpha (\alpha - 1)} > 0 \) and normalizing \( b_{n+1,i} \), (13) is obtained.
3. Empirical Results
Fifteen publicly listed Malaysian companies are randomly selected for the empirical study. These companies are divided into 5 data sets each comprising of five-stock portfolios where each data set consists of at least 3 different sectors of industry. This diversification aims to reduce investment risks. The details of the selected companies are listed in Table 1. The period of trading of the stocks is from 1st March 2006 until 2nd August 2012, consisting of a total of 1500 trading days.

| Data Set | Malaysian Companies in Each Portfolio |
|----------|---------------------------------------|
| D        | IOI Corporation, Carlsberg Brewery Malaysia, British American Tobacco, Nestle, Digi.com |
| E        | Public Bank, Kulim, KLCC Property Holdings, AEON Corporation, Kuala Lumpur Kepong |
| F        | AMMB Holdings, Berjaya Sports TOTO, Air Asia, Gamuda, Genting |
| G        | AEON Corporation, British American Tobacco, Kulim, Nestle, Digi.com |
| H        | Digi.com, Public Bank, KLCC Property Holdings, Carlsberg Brewery Malaysia, Kuala Lumpur Kepong |

The binomial approximation of the Bregman universal portfolio generated by \( f(t) = \pm t^\alpha - \alpha t \) is given by (13). This portfolio is run over the selected data sets D, E, F, G and H. The accumulated wealth \( S_{1500} \) after 1500 trading days together with the final portfolios \( b_{1501} \) are listed in Table 2 for selected values of parameters \( \delta \) and \( \alpha \). Among these five portfolios, data set E is a good portfolio achieving 8.2725 units in return. It is observed that the wealth achieved for data sets G and H are 4.778 and 5.06 units respectively, exhibiting an average performance. Portfolios D and F on the other hand perform poorly for the achieved 2.56 and 1.36 respectively.

Table 2 reveals that there is some pattern in these portfolios. For data sets D, E and F, it is observed that as the value of \( \delta \) increases, the corresponding wealth \( S_{1500} \) also increases and performs best when parameter \( \alpha \) remains at 3. Figure 1 shows that in order to obtain a higher return it is necessary to increase the value of \( \delta \). However, the value of \( \alpha \) will remain around 3. Whereas, portfolios G and H are found to achieve higher wealth \( S_{1500} \) for parameter \( \alpha \) close to 1, for all range of values of \( \delta \). In this study, should the parameter \( \alpha \)
value remain around 3 or 1, the parameter $\delta$ reacts differently on different data set exhibiting the best performance locally.

![Figure 1](image)

**Figure 1.** Scatter plot of $S_{1500}$ against $\delta$ and the portfolio proportions $b_i$ against $\delta$ at $\alpha = 3$ for data set D.

A comparison of the empirical results with the reverse Helmbold universal portfolio from Tan, Kuang and Lee [22] is shown in Table 3. The wealth achieved for data sets D, E and H are higher for the Bregman universal portfolio. On the other hand, data sets F and G seem to perform better for the reverse Helmbold portfolio. Bregman portfolio can perform better than the reverse Helmbold portfolio on certain data sets and not on all data sets. In general, the performance of a universal portfolio depends on the performance of the constituent stocks in the portfolio.

| Set | Bregman $S_{1500}$ | Reverse Helmbold $S_{1500}$ |
|-----|---------------------|-----------------------------|
| D   | 2.56                | 2.52452                     |
| E   | 8.2725              | 8.01922                     |
| F   | 1.36                | 1.43580                     |
| G   | 4.778               | 5.71007                     |
| H   | 5.06                | 5.01154                     |

**Table 3.** The comparison of wealth obtained after 1500 trading days by Bregman universal portfolio and the reverse Helmbold universal portfolio over the data sets D, E, F, G and H.

The ability of the universal portfolio to assign proper weights to the constituent stocks to achieve higher investment returns is demonstrated for portfolios E, G and H in this empirical study. This study suggests that the companies in portfolio E perform best in a group and also help to maximize return while collaborating with companies from portfolios D and F, as presented in portfolios G and H. Higher returns can be achieved over a longer period of trading of the stocks. The binomial approximation of the Bregman universal portfolio introduced here contributes to the inventory of universal portfolios available for investment.
In conclusion, the algorithm does not perform well if there are weak stocks during the trading period. The algorithm is weak if there are weak stocks in the portfolio. Under the right conditions and if the portfolio contains some blue-chip stocks, the algorithm enunciated here is expected to perform well. The conditions include no downturn in the local economy and economic stability in the global market. The advantages of using a universal portfolio is that no assumption is made on the stochastic model of the stock prices, because it is difficult to verify the mathematical model.

References
[1] Markowitz H 1952 The Journal of Finance, 7(1), 77 – 91
[2] Sharpe WF 1963 Management Science, 9(2), 277 – 293
[3] Bernoulli D 1954 Econometrica: Journal of the Econometric Society, 23 – 36
[4] Kelly JL 1956 Bell System Technical Journal, 35(4), 917 – 926
[5] Latane HA 1959 Journal of Political Economy, 67(2), 144 – 155
[6] Breiman L 1961 Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, 1, 65 – 78
[7] Thorp EO 1969 Revue de l’Institut International de Statistique, 37(3), 273 – 293
[8] Thorp EO 1975 Stochastic Optimization Models in Finance, 599 – 619
[9] Markowitz HM 1976 The Journal of Finance, 31(5), 1273 – 1286
[10] Bell RM and Cover TM 1980 Mathematics of Operations Research, 5(2), 161 – 166
[11] Cover TM 1984 IEEE Transactions on Information Theory, 30(2), 369 – 373
[12] Cover TM and Gluss DH 1986 Advances in Applied Mathematics, 7(2), 170 – 181
[13] Baron AR and Cover TM 1988 IEEE Transaction on Information Theory, 34(5), 1097 – 1100
[14] Algoet PH and Cover TM 1988 The Annals of Probability, 16(2), 876 – 898
[15] Algoet PH 1992 The Annals of Probability, 20(2), 901 – 941
[16] Cover TM 1991 Mathematical Finance, 1(1), 1 – 29
[17] Cover TM and Ordentlich E 1996 IEEE Transactions on Information Theory, 42(2), 348 – 363
[18] Helmbold DP, Shapire RE, Singer Y and Warmuth MK 1998 Math. Finance, 8, 325 – 347
[19] Tan CP and Tang SF 2003 Malaysian Journal of Science, 22(2), 127 – 133
[20] Tan CP and Lim WX 2011 Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering and Medicine, 2(6), 158 – 165
[21] Basu A, Shioya H and Park C 2011 Statistical Inference: The Minimum Distance Approach. Chapman and Hall, Boca Raton FL
[22] Tan CP, Kuang KS and Lee YJ 2017 AIP Conference Proceedings 1830, 020023
[23] Tan CP and Kuang KS 2015 AIP Conference Proceedings 1691, 040025
[24] Tan CP and Kuang KS 2017 AIP Conference Proceedings 1830, 020021