Quantitative Finance, 2017
Vol. 17, No. 5, 763–780, http://dx.doi.org/10.1080/14697688.2016.1223337

Binary switch portfolio

TENGFEI LI†‡, KANI CHEN§, YANG FENG*¶ and ZHILIANG YING¶‡

†Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
‡Shanghai Center for Mathematical Sciences, Shanghai, China
§Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong
¶Department of Statistics, Columbia University, New York, NY, USA

(*Received 17 May 2014; accepted 5 August 2016; published online 4 November 2016*)

We propose herein a new portfolio selection method that switches between two distinct asset allocation strategies. An important component is a carefully designed adaptive switching rule, which is based on a machine learning algorithm. It is shown that using this adaptive switching strategy, the combined wealth of the new approach is a weighted average of that of the successive constant rebalanced portfolio and that of the 1/N portfolio. In particular, it is asymptotically superior to the 1/N portfolio under mild conditions in the long run. Applications to real data show that both the returns and the Sharpe ratios of the proposed binary switch portfolio are the best among several popular competing methods over varying time horizons and stock pools.

Keywords: Aggregating algorithm; Asset return; Bayesian analysis; Portfolio selection; Supervised learning; Universal portfolio

JEL Classification: C11, C38, C44, G11

1. Introduction

Portfolio selection is an important subject in economics and finance. Among a variety of methods proposed in academia and implemented in practice, Markowitz's mean–variance portfolio (Markowitz 1952) has become an industry standard and Cover's Universal Portfolio (CUP) (Cover 1991) is one of the most competitive methods. Both methods have been further extended and improved in the past decades. On the other hand, the naive 1/N rule, which invests equally in each of the N assets, serves as a benchmark for evaluating portfolio performance. Empirical studies support that this naive 1/N rule outperforms many sophisticated portfolios; see, for example, DeMiguel *et al*. (2009). And it has been shown (Pflug *et al*. 2012) that the 1/N investment strategy is optimal under high model ambiguity, which means, for a broad class of risk measures, as the uncertainty concerning the probabilistic model increases, the optimal decisions tend to the uniform (1/N) investment strategy.

Markowitz's mean–variance portfolio, despite its popularity, relies on the assumption about the statistical stationarity of the asset returns as well as the knowledge of the underlying distribution (mean and covariance matrix). In practice, one first estimates the mean and covariance from historical data and then plugs in the resulting estimators. It has been found that Markowitz's portfolio usually underperforms the naive 1/N rule, both in simulations and in practical implementations (DeMiguel *et al*. 2009, Tu and Zhou 2011). On the other hand, Cover (1991) showed that his universal portfolio approach has the advantage of not relying on statistical assumptions of the underlying returns and thus can be implemented without parameter estimation. He demonstrated that the CUP is very stable in terms of its performance and the method is universally applicable. Further extensions of the CUP can be found in Jamshidian (1992), Cover (1996), Ordentlich and Cover (1996), Cover and Ordentlich (1996), Kalai and Vempala (2003) and Cross and Barron (2003), among many others. However, it is worth noting that the 1/N method can outperform the CUP in many situations (Belentepe and Wyner 2005).

In the area of online prediction in computer science, the aggregating algorithm (AA) proposed by Vovk (1990) generalizes the weighted majority algorithm (Littlestone and Warmuth 1989) to a much wider class of outcome and action spaces. Vovk and Watkins (1998) applied the AA to portfolio selection. That application can be shown (see section 2) to result in a generalization of the CUP that includes an extra learning rate parameter η, with the CUP being just a special case with $\eta = 1$. Ignoring the cases with $\eta > 1$, they focused on the cases with $\eta \leq 1$ and obtained a bound on the extra loss, as compared with the best constant rebalanced portfolio (BCRP) (Belentepe and Wyner 2005). Furthermore, Gaivoronski and

*Corresponding author. Email: yangfeng@stat.columbia.edu

Stella (2000) proposed what they called the successive constant rebalanced portfolio (SCRP) method, which is a typical follow-the-leader rule. We note that the AA approximates the SCRP when the learning rate η is large (approaches infinity). In summary, the AA method contains, as special cases, the 1/N, the CUP and the SCRP with η being equal to 0, 1 and $+\infty$, respectively.

In this paper, we propose a method called binary switch portfolio (BSP). We first investigate how to adaptively make use of the AA with different values of η. We found that switching between the 1/N rule, which is the most conservative constant rebalanced portfolio (CRP), and the SCRP, which is the most aggressive one, provides a simple and approximately optimal investment strategy. A key ingredient in the BSP is the determination of when to switch between the two opposing strategies. Our approach is to apply machine learning techniques to develop a stable switching mechanism. Applications of such machine learning techniques to the description of market regimes and prediction of future returns have given rise to a broad interest; Huang *et al*. (2005), Qian and Rasheed (2006) and Patel *et al*. (2015).

Because the SCRP adopts the follow-the-leader rule and therefore can be viewed as momentum-based strategy, while the 1/N rule seeks to increase (decrease) the holdings of those stocks that perform relatively worse (better) than the remaining ones and therefore can be viewed as mean reversion-oriented, the BSP may be interpreted as an investment strategy that alternates between the momentum and mean reverting trading strategies. The intuition for the existence of separate blocks of securities and time ranges in which either a momentum-based strategy or a mean reversion-based one performs better can also be found in Hwang and Rubesam (2015). We provide here further evidence that such blocks not only exist, but also can be identified via machine learning techniques.

This paper provides theoretical justification as well as empirical evidence for the proposed BSP strategy. It first applies the BSP to a data-set that contains 59 stocks from the Hong Kong Stock Exchange from 2008 to 2013. For comparison with existing methods, two additional data-sets, one containing the stocks from the New York Stock Exchange from 1962 to 1984 (Cover 1991), and the other containing the Fama–French 30 industries portfolios (Fama and French 1997), are also analysed. The results show that the BSP significantly outperforms the AA with a fixed learning rate, including the 1/N rule, the CUP and the SCRP, under various time horizons.

The rest of the paper is organized as follows. Section 2 provides a technical background by reviewing the AA approach and establishing a bound on the extra loss of the AA as compared with the BCRP under any fixed learning rate. In section 3, the BSP is introduced and its properties studied. The BSP is applied to several data-sets in section 4 where comparisons with other methods are also given. Some concluding remarks are given in section 5. All the regularity conditions and technical proofs are relegated to Appendices 1 and 2.

2. The AA approach

2.1. Commonly used portfolios

We are concerned with a market containing p stocks in which discrete time points $t_0 < t_1 < \cdots < t_n < \cdots$ are under consideration. Let $s_{n,j}$ be the price of the jth stock at time t_n. We assume that there is no dividend or transaction cost and that short selling is not allowed. Define $x_{n,j} = s_{n,j}/s_{n-1,j}$, the simple gross return (price relative) of stock j for the nth time period $(t_{n-1}, t_n]$. Let $\boldsymbol{x}_n = (x_{n,1}, \ldots, x_{n,p})^T$ be the corresponding vector. Here and in the sequel, $\boldsymbol{x}^T$ denotes the transpose of a vector $\boldsymbol{x}$. A portfolio corresponds to a p-vector $\boldsymbol{b} = (b_1, \ldots, b_p)^T$, where $b_j \geq 0$ $(j = 1, \ldots, p)$ and $\sum_j b_j = 1$. The set of such portfolios is denoted by B. Let $\boldsymbol{b}_n = (b_{n,1}, b_{n,2}, \ldots, b_{n,p})^T$ be the portfolio used in the nth time period $(t_{n-1}, t_n]$. Thus, $\mathbf{B}_n = (\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)$ can be viewed as an investment strategy (portfolio sequence). Furthermore, let W_n be the total (cumulative) wealth at time t_n and assume, for simplicity, $W_0 = 1$ (unit initial wealth). Therefore, the cumulative wealth under strategy $\mathbf{B}_n$ is

$$W_n = W_{n-1}\boldsymbol{x}_n^T\boldsymbol{b}_n. \tag{1}$$

The portfolio construction has to be non-anticipating, meaning that it cannot use information from future stock prices. Let $\mathcal{F}_n = \sigma\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ be the σ-field generated from random variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, representing the information accumulation up to time t_n. Then, $\boldsymbol{b}_n \in \mathcal{F}_{n-1}$, meaning that $\boldsymbol{b}_n$ is $\mathcal{F}_{n-1}$ measurable (adapted to $\mathcal{F}_{n-1}$).

A particularly simple class of investment portfolios is the so-called CRP; see Cover (1991). For such portfolios, allocation proportions remain the same over the entire investment horizon, i.e. $\boldsymbol{b}_1 = \boldsymbol{b}_2 = \cdots = \boldsymbol{b}_n$. A special case is the naive and widely used '1/N portfolio', also known as uniformly constant rebalanced portfolio, which takes form $b_{n,1} = \cdots = b_{n,p} = 1/p$. We shall denote the 1/N portfolio by $\boldsymbol{b}^{(0)}$, i.e.

$$\boldsymbol{b}^{(0)} = \left(\frac{1}{p}, \frac{1}{p}, \ldots, \frac{1}{p}\right)^T.$$

In general, for a CRP $\boldsymbol{b}$, let

$$W_n(\boldsymbol{b}) = \prod_{i=1}^{n} \boldsymbol{x}_i^T\boldsymbol{b}$$

represent the total wealth at t_n. Let

$$\boldsymbol{b}_n^* = \arg\max_{\boldsymbol{b}\in B} W_n(\boldsymbol{b}).$$

Since $W_n(\boldsymbol{b}) \leq W_n(\boldsymbol{b}_n^*)$, it follows that $\boldsymbol{b}_n^*$ would be the BCRP if $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ were known in advance. We note that the BCRP is not implementable in practice since the $\boldsymbol{x}_i$ are not known in advance and it serves only as a benchmark for purpose of comparison.

Cover (1991) proposed what he called the universal portfolio and showed that such a portfolio achieves asymptotically the same growth rate as that of the BCRP. CUP is defined as follows:

$$\boldsymbol{b}_n^{(1)} = \frac{\int \boldsymbol{b}W_{n-1}(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}{\int W_{n-1}(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}, \tag{2}$$

where π is a pre-specified prior distribution on B. A commonly used prior, to be denoted here by π_0, is the uniform distribution on B.

The AA was introduced in the context of prediction strategies by Vovk (1990) for general outcome and action spaces. Using the AA, a general class of portfolios has the form

$$\boldsymbol{b}_n^{(\eta)} = \frac{\int \boldsymbol{b}[W_{n-1}(\boldsymbol{b})]^\eta \pi(\mathrm{d}\boldsymbol{b})}{\int [W_{n-1}(\boldsymbol{b})]^\eta \pi(\mathrm{d}\boldsymbol{b})}, \tag{3}$$

where $\eta \geq 0$ is a pre-chosen constant. In general, η may be viewed as the 'learning rate' of the AA: a larger value of η implies that among all portfolios, we put more weight on those with better historical performances, and less weight on those with poorer performances. As such, it can be understood as a measure of the momentum level. In particular, the special case of $\eta = 1$ corresponds to the CUP, whereas that of $\eta = 0$ corresponds to the 1/N rule when π is the uniform distribution.

To compare performances among different portfolios, we need to consider the corresponding cumulative wealths and, in particular, their growth rates. Denote by $W_n^{(\eta)}$ the cumulative wealth at t_n achieved by the AA with learning rate η. Cover (1991) compared the log-wealth of the CUP, $\log W_n^{(1)}$, with the log-wealth of the BCRP, $\log W_n(\boldsymbol{b}_n^*)$, and showed that they are of the same order. Vovk and Watkins (1998) further proved that, when the prior π has a finite or countable support,

$$\log W_n^{(\eta)} \geq \log[W_n(\boldsymbol{b}_n^*)] - \frac{(p-1)\log n}{2\eta}(1+o(1)), \quad \text{for any } \eta \in (0,1]. \tag{4}$$

On the other hand, we know that $\log W_n^{(\eta)} \leq \log[W_n(\boldsymbol{b}_n^*)]$ by the definition of $\boldsymbol{b}_n^*$. Since $\log W_n^{(\eta)}$ is of order n, it is seen from (4) that as η gets larger, the performance of the AA is closer to that of the BCRP, i.e. the bound becomes tighter. However, (4) does not hold for $\eta > 1$. In the next subsection, a systematic study will be conducted for the case with $\eta > 1$.

2.2. A general bound for AA

In this subsection, we develop a general inequality that covers all positive values of η and also extends (4). First, we need the following regularity conditions.

(A.1) There exists a constant $\delta > 0$, such that

$$\underline{\lim}_n \lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^n \frac{\boldsymbol{x}_i\boldsymbol{x}_i^T}{\boldsymbol{x}_i^T\boldsymbol{x}_i}\right) > \delta,$$

where $\lambda_{\min}(M)$ ($\lambda_{\max}(M)$) of matrix M denotes its minimum (maximum) eigenvalue.

(A.2) There exist $0 < c_1 < c_2$, such that, for $i = 1, 2, \ldots, n$, each element of the vector $\boldsymbol{x}_i$ falls in $[c_1, c_2]$.

(A.3) The density of the prior π, denoted by f, exists and is bounded away from 0, i.e. $\inf_B f(\boldsymbol{b}) > 0$.

Conditions (A.1) and (A.2) are also assumed in Gaivoronski and Stella (2000). Note that the trace of the non-negative definite matrix in Condition (A.1), which always dominates its maximum eigenvalue, is always 1. Thus, Condition (A.1) implies that the matrix is well conditioned. Condition (A.2) requires that the values of gross returns are bounded away from 0 and ∞, i.e. price fluctuations of stocks cannot be extreme. Condition (A.3) is a technical assumption, which is common for a prior on a compact support.

The following lemma quantifies the difference between the AA and the BCRP in terms of their corresponding asset allocation vector at time $n+1$ and n, respectively. Its proof will be provided in appendix 2.

Lemma 2.1 *Suppose Conditions* (A.1)–(A.3) *are satisfied. Then, for any given constant $\eta_0 > 0$, there exists $N > 0$ such that, for all $n > N$ and $\eta \in [\eta_0, +\infty)$,*

$$\|\boldsymbol{b}_{n+1}^{(\eta)} - \boldsymbol{b}_n^*\| \leq \sqrt{\frac{2p}{\delta\eta}\frac{\log n}{n}}.$$

Remark 1 From lemma 2.1, for a fixed η, the asset allocation vector of the AA at time $n+1$ is close to that of the BCRP at time n in L_2 norm when the time period is large. On the other hand, for a fixed n, the conclusion also holds as the learning rate $\eta \to \infty$.

From lemma 2.1, we immediately have the following proposition, showing that the SCRP (Gaivoronski and Stella 2000), is the limiting case of the AA with $\eta = \infty$. As a result, we use the notation $\boldsymbol{b}_n^{(\infty)}$ to represent the portfolio of the SCRP.

Proposition 2.2 *Under Conditions* (A.1)–(A.3)*, there exists $N > 0$ such that, for all $n > N$,*

$$\lim_{\eta\to\infty} \boldsymbol{b}_n^{(\eta)} = \boldsymbol{b}_n^{(\infty)}.$$

Using lemma 2.1, the following theorem provides a lower bound of the cumulative wealth $W_n^{(\eta)}$ for $\eta \geq \eta_0$, with the proof given in appendix 2.

Theorem 2.3 *Under Conditions* (A.1)–(A.3)*, for any given $\eta_0 > 0$, there exists $N > 0$ such that, for all $n > N$ and $\eta \in [\eta_0, +\infty)$,*

$$\log W_n^{(\eta)} \geq \log[W_n(\boldsymbol{b}_n^*)] - \min\left\{\left(\frac{2\sqrt{2}c_2 p}{c_1\sqrt{\delta\eta}}\sqrt{n\log n} + \frac{2c_2^2 p}{c_1^2\delta}\log n\right), \frac{(p-1)\log n}{2\eta \mathrm{I}_{\{\eta\leq 1\}}}\right\}, \tag{5}$$

where c_1 and c_2 are those defined in Condition (A.2)*, and $\mathrm{I}_{\{\eta\leq 1\}} = 1$ if $\eta \leq 1$ and 0 otherwise.*

Remark 2 Theorem 2.3 generalizes the results in Vovk and Watkins (1998) for $\eta \leq 1$, Gaivoronski and Stella (2000) for $\eta \to \infty$ and extends to the case of $\eta \in (1, +\infty)$.

Remark 3 When $\eta \to \infty$, (5) reduces to

$$\log W_n^{(\infty)} \geq \log[W_n(\boldsymbol{b}_n^*)] - \frac{2c_2^2 p \log(n-1)}{c_1^2\delta}(1+o(1)),$$

which was proved in Gaivoronski and Stella (2000).

Remark 4 The AA defines a class of portfolios indexed by the learning rate η. The naive 1/N rule, the CUP and the SCRP correspond to the AA with $\eta = 0$ (when $\pi = \pi_0$), $\eta = 1$ and $\eta = \infty$, respectively. Except for the 1/N rule, they have the following asymptotic property,

$$\frac{1}{n}[\log W_n^{(\eta)} - \log W_n(\boldsymbol{b}_n^*)] \to 0, \tag{6}$$

as $n \to \infty$.

3. The switching method

In this section, we first study the behaviour of the oracle learning rate of the AA, and then propose a switching method, whose asymptotic behaviour is delineated.

3.1. Oracle learning rate of the AA

At each time point t_n, there exists a learning rate η_n^*, which maximizes the returns during $[t_{n-1}, t_n)$, $n = 1, 2, \ldots$. Specifically,

$$\eta_n^* = \arg \max_{\eta \in [0,+\infty]} y_n(\eta), \text{ where } y_n(\eta) = \boldsymbol{x}_n^T \boldsymbol{b}_n^{(\eta)}, n \geq 2, \text{ and } \eta_1^* = 0. \tag{7}$$

We call η_n^* the 'oracle learning rate' as it leads to the maximum return if one were given the information during $[t_{n-1}, t_n)$, which is unknown at time t_{n-1}. As a result, it is desirable to develop a strategy that can approximate this 'oracle learning rate'.

We recognize that η_n^* is not adapted to $\mathcal{F}_{n-1}$, i.e. it is not known before t_n. If it were known for $n = 1, 2, \ldots$, then the cumulative wealth W_n^* would then be achieved by the strategy

$$\mathbf{B}_n^* = (\boldsymbol{b}_1^{(\eta_1^*)}, \boldsymbol{b}_2^{(\eta_2^*)}, \ldots, \boldsymbol{b}_n^{(\eta_n^*)}).$$

Next we search for the η_n^* in $[t_{n-1}, t_n)$, $n = 1, 2, \ldots$, to get the cumulative wealth as close to W_n^* as possible. This W_n^*, instead of $W_n(\boldsymbol{b}_n^*)$, is our new benchmark in this section. It is larger than $W_n^{(\eta)}$ for any fixed η. Actually, it turns out that it is far larger than $W_n(\boldsymbol{b}_n^*)$, the benchmark we used in the last section, in our numerical experience.

The following list is a segment of η_n^* in 1974 based on daily return data-set of *Arcos Dorados Holdings Inc*. and *Ashford Hospitality Prime Inc*. in 1962–1984; see the data-set of New York Stock Exchange analysed in section 4. We optimize η that ranges among $\eta \in A = \{0, \eta_{1,1}, \eta_{1,2}, \ldots, \eta_{1,100}, \eta_{2,1}, \eta_{2,2}, \ldots, \eta_{2,400}\}$, where $\eta_{1,i}, i = 1, 2, \ldots, 100$ is a geometric sequence with $\eta_{1,1} = .0001, \eta_{1,100} = .9999$, and $\eta_{2,i}, i = 1, 2, \ldots, 400$ is an equal-spaced sequence with $\eta_{2,1} = 1$, $\eta_{2,400} = 5000$. Here, η_n^* is calculated as follows. We first evaluate $\boldsymbol{b}_n^{(\eta)}$ using (3) for each $n = 1, 2, \ldots$, and each η in set A, where the set A is chosen to ensure the search area is dense in both $[0, 1]$ and $[1, C]$, where C is sufficiently large. Then we set $\eta_n^* = \arg\max_{\eta \in A} y_n(\eta)$ where $y_n(\eta)$ is defined in (7). Note that if $\eta_n^* = C$, we set it as ∞. The list is

$$\begin{gathered} 0\ 0\ 0\ 0\ 0\ 0\ 0\ 641\ 0\ 561\ 561\ 0\ \infty\ \infty\ \infty\ \infty\ \infty\ \infty\ 0\ 0\ \infty\ \infty\ 0\ 0\ 0\ 0 \\ \infty\ 0\ 0\ \infty\ \infty\ \infty\ 0\ \infty\ 0\ 0\ 61\ 0\ \infty\ \infty\ 61\ 0\ 0\ \infty\ 0\ 0\ \infty\ 0\ \infty\ 0\ 0\ 21\ 0 \end{gathered} \tag{8}$$

We observe in (8) that 0 and ∞ take a large proportion among all possible η_n^*, $n = 1, 2, \ldots$. In fact, the rate of $\eta_n^* = 0$ or ∞ is 84%.

The following lemma characterizes the derivative $\partial \log[y_n(\eta)]/\partial\eta$, which will be used in Theorem 3.2 to quantify the oracle learning rate η_n^* under different scenarios.

LEMMA 3.1 *Let $\mathfrak{b}_{n-1}^\eta$ be a random variable that induces the probability $\mathrm{P}_{n,\eta}$ on B. Then,*

$$\frac{\partial \log[y_n(\eta)]}{\partial \eta} = \mathrm{Cov}_\eta\big(\boldsymbol{x}_n^T \mathfrak{b}_{n-1}^\eta, \log W_{n-1}(\mathfrak{b}_{n-1}^\eta)\big)/\mathrm{E}_\eta(\boldsymbol{x}_n^T \mathfrak{b}_{n-1}^\eta),$$

where Cov_η and E_η are covariance and expectation with respect to $\mathfrak{b}_{n-1}^\eta$, conditioning on all $\boldsymbol{x}_i$, $i = 1, 2, \ldots$

We note that $\mathfrak{b}_n^\eta$ and $\boldsymbol{b}_n^{(\eta)}$ are different: the former is Random, while the latter is its expectation, i.e. $\boldsymbol{b}_n^{(\eta)} = \mathrm{E}_\eta \mathfrak{b}_{n-1}^\eta$. Using lemma 3.1 we have the following theorem.

THEOREM 3.2 *Suppose that the prior $\pi = \pi_0$.*

(a) *In general, we have*

$$\begin{cases} \eta_n^* = 0, & \text{if } \mathrm{Cov}_\eta\big(\boldsymbol{x}_n^T \mathfrak{b}_{n-1}^\eta, \log W_{n-1}(\mathfrak{b}_{n-1}^\eta)\big) < 0 \text{ for all } \eta \in (0, +\infty); \\ \eta_n^* = +\infty, & \text{if } \mathrm{Cov}_\eta\big(\boldsymbol{x}_n^T \mathfrak{b}_{n-1}^\eta, \log W_{n-1}(\mathfrak{b}_{n-1}^\eta)\big) > 0 \text{ for all } \eta \in (0, +\infty). \end{cases}$$

(b) *Suppose that Conditions* (A.1), (A.2)*, and* (A.4)–(A.6) *in appendix A hold. Then, for any constant $\alpha > 1$ and any sequence $\eta_{1,n}$ satisfying*

$$\frac{n}{(\log n)^\alpha} \eta_{1,n} \longrightarrow \infty,$$

there exists a constant $N > 0$ such that, for any $n > N$,

$$\eta_n^* \in [0, \eta_{1,n}] \cup \{+\infty\}.$$

In addition , if Condition (A.7) *also holds, then*

$$\eta_n^* \in [\eta_{2,n}, \eta_{1,n}] \cup \{0, +\infty\}$$

for all large n provided $n\eta_{2,n} \to 0$.

Remark 5 The above theorem shows us the range of possible values for η_n^*. Part (a) provides a sufficient condition for $\eta_n^* \in \{0, \infty\}$ and it follows directly from lemma 3.1. An intuitive explanation is that the covariance term is a measure of agreement between the future returns $\boldsymbol{x}_n^T \boldsymbol{b}$ and the historical returns $\log W_{n-1}(\boldsymbol{b})$. If the covariance term is always positive, then it indicates that we should choose the SCRP. On the other hand, if the term is always negative, it implies that one may need to choose the more conservative 1/N portfolio. In particular, suppose the prior $\pi(\mathrm{d}\boldsymbol{b})$ has a discrete support. For example, suppose $\pi(\boldsymbol{b}_1) + \pi(\boldsymbol{b}_2) = 1$ with positive $\pi(\boldsymbol{b}_1)$ and $\pi(\boldsymbol{b}_2)$. Then, we have

$$\begin{aligned} &\mathrm{Cov}_\eta\big(\boldsymbol{x}_n^T \mathfrak{b}_{n-1}^\eta, \log W_{n-1}(\mathfrak{b}_{n-1}^\eta)\big) \\ &\propto \left[\int \boldsymbol{x}_n^T \boldsymbol{b} \log W_{n-1}(\boldsymbol{b}) W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})\right]\left[\int W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})\right] \\ &- \left[\int \boldsymbol{x}_n^T \boldsymbol{b} W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})\right]\left[\int \log W_{n-1}(\boldsymbol{b}) W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})\right] \\ &\propto (\boldsymbol{x}_n^T \boldsymbol{b}_1 - \boldsymbol{x}_n^T \boldsymbol{b}_2)(\log W_{n-1}(\boldsymbol{b}_1) - \log W_{n-1}(\boldsymbol{b}_2)). \end{aligned} \tag{9}$$

Remark 6 Part (b) tells us that, under some regularity conditions, $[\eta_{2,n}, \eta_{1,n}] \cup \{0, \infty\}$ is the area to search for η_n^* asymptotically, where $1/n \ll \eta_{2,n} < \eta_{1,n} \ll (\log n)^\alpha/n$. Besides Conditions (A.1)–(A.3) that are explained following lemma 2.1, (A.4) can be found in lemma 4.1 of Cover (1991). Condition (A.5) imposes a limit on the growth rate of the optimal portfolio and this condition is generally satisfied. To see this, the optimal portfolio is obviously better than the risk-free asset, which has an exponential growth rate (positive interest rate). On the other hand, in the long run, the average log-return of any stock should be bounded from above. As a result, limiting the log-return to an interval is reasonable. Condition (A.6) is a technical condition which should be satisfied in most situations. In addition, it can be relaxed by taking a higher order term in the Taylor series expansion in the proof, details of which are omitted. Condition (A.7) assumes that $\boldsymbol{x}_n\boldsymbol{b}$ and $\log W_{n-1}(\boldsymbol{b})$ are correlated in terms of non-vanishing correlation when $n \to \infty$, as $\boldsymbol{b}$ is random with distribution $\pi(\boldsymbol{b})$.

It is also seen in (8) that usually there is a momentum for the choice of η_n^*, or a coherence of the η_n^* and its neighbours η_{n-1}^* and η_{n+1}^*. This motives us to derive a data-driven method of choosing η_n to approximate η_n^* in the next subsection.

3.2. The BSP

From theorem 3.2(b), one only needs to search the area $\{0, \infty\} \cup [\eta_{2,n}, \eta_{1,n}]$ for the optimal η. Note that if $\eta_{1,n} \to 0$, any fixed learning rate in $(0, \infty)$ will be suboptimal in the long-time horizon. To enhance stability and expedite computation, we propose to restrict our search to only 0 and ∞, which we would call a 'binary switch' algorithm. Throughout the subsection, we redefine

$$\eta_n^* = \arg \max_{\eta \in \{0,+\infty\}} y_n(\eta) \text{ and } W_n^* = W_{n-1}^* \boldsymbol{x}_n^T \boldsymbol{b}_n^{(\eta_n^*)}, \quad n = 1, 2, \ldots. \tag{10}$$

Define

$$\mathcal{C}_1 = \left\{n \geq 2 | \eta_n^* = \eta_{n-1}^*\right\} \text{ and } \mathcal{C}_2 = \left\{n \geq 2 | \eta_n^* \neq \eta_{n-1}^*\right\}. \tag{11}$$

Since η_{n-1}^* is available at time t_n, we will have an accurate estimator of η_n^* as long as the unknown sets $\mathcal{C}_1$ and $\mathcal{C}_2$ can be precisely predicted. Now suppose a random function κ is a classifier defined on $\mathbb{N}_+ \backslash \{1\}$ that satisfies

$$\kappa : \mathbb{N}_+ \backslash \{1\} \mapsto \{1, 2\}, \text{ and } \kappa(n) \in \mathcal{F}_{n-1}, \quad \text{for any } n \in \mathbb{N}_+ \backslash \{1\}. \tag{12}$$

Here, the value of $\kappa(n) = i$ indicates that we estimate $n \in \mathcal{C}_i$, for $i = 1, 2$. Based on the fact that there is a coherence between η_n^* and its neighbours, a binary switch procedure is to be specified in the following. Here, we use $\#A$ to denote the cardinality of the set A.

Definition 3.3 Let

$$\mathcal{E}_n(\kappa) = \#\{k = 3, 4, \ldots, n | \kappa(k) = 2, k \in \mathcal{C}_1 \text{ or } \kappa(k) = 1, k \in \mathcal{C}_2\}/(n-2), \tag{13}$$

which is the empirical classification error up to time t_n. Also, let

$$\tilde{\eta}_n = \begin{cases} \eta_{n-1}^*, & \text{if } \kappa(n) = 1 \text{ and } \mathcal{E}_{n-1}(\kappa) \leq u; \\ 0, & \text{otherwise,} \end{cases}$$

where u is a pre-specified positive threshold, and η_n^* is defined in (10). We call the AA with learning rate $\tilde{\eta}_n$ BSP. In addition, $\tilde{\mathbf{B}}_n = (\boldsymbol{b}_1^{(0)}, \boldsymbol{b}_2^{(\tilde{\eta}_2)}, \boldsymbol{b}_3^{(\tilde{\eta}_3)}, \ldots, \boldsymbol{b}_n^{(\tilde{\eta}_n)})$ is called the binary switch strategy.

Remark 7 The BSP is a conservative procedure in the sense that it will follow the 1/N unless it is confident about the assignments and $\eta_{n-1}^* = \infty$.

Remark 8 Although risk is not being explicitly considered in the CUP and its extensions, there is a strong connection between CUP and the Markowitz portfolio, which balances the return and risk, as being pointed out in Belentepe and Wyner (2005). In fact, by the Taylor series expansion, when $x = o_p(1)$,

$$\mathrm{E}[\log(1+x)] \approx \mathrm{E}(x) - \frac{1}{2}\mathrm{E}[(x - \mathrm{E}(x))^2],$$

which can be viewed as a mean-variance rule or the well-known Kelly's rule (Kelly 1956). Consequently, the risk is also controlled in the CUP and its variants including BSP.

To study the theoretical properties of the BSP, we define the concept of compressed cumulative wealth.

Definition 3.4 Let $\{\mathcal{G}_i, i \geq 1\}$ be a filtration. The corresponding $\mathcal{G}_n-$compressed cumulative wealth achieved by strategies $\mathbf{B}_n = (\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)$ is defined by

$$\exp\left\{\sum_{i=1}^{n} \mathrm{E}[\log \boldsymbol{x}_i^T \boldsymbol{b}_i | \mathcal{F}_{i-1}, \mathcal{G}_i]\right\}.$$

In particular, we will use $W_{n|\mathcal{G}_n}^{(0)}$, $\tilde{W}_{n|\mathcal{G}_n}$, and $W_{n|\mathcal{G}_n}^*$ to denote the $\mathcal{G}_n-$compressed cumulative wealth using strategies $\mathbf{B}_n^{(0)}$ (1/N) , $\tilde{\mathbf{B}}_n$ (BSP), and $\mathbf{B}_n^*$ (oracle), respectively.

Remark 9 It is obvious that

$$W_{n|\mathcal{F}_n}^{(0)} = W_n^{(0)}, W_{n|\mathcal{F}_n}^* = W_n^*.$$

The compressed cumulative wealth compresses the information of returns by conditional expectation on a filtration smaller than $\mathcal{F}_n$.

Define

$$\mathcal{C}_{0,0} = \{k = 2, 3, \ldots, n | \eta_{k-1}^* = 0, \eta_k^* = 0\},$$
$$\mathcal{C}_{\infty,\infty} = \{k = 2, 3, \ldots, n | \eta_{k-1}^* = +\infty, \eta_k^* = +\infty\},$$
$$\mathcal{C}_{0,\infty} = \{k = 2, 3, \ldots, n | \eta_{k-1}^* = 0, \eta_k^* = +\infty\},$$
$$\mathcal{C}_{\infty,0} = \{k = 2, 3, \ldots, n | \eta_{k-1}^* = +\infty, \eta_k^* = 0\}.$$

Then, $\mathcal{C}_1 = \mathcal{C}_{0,0} \cup \mathcal{C}_{\infty,\infty}$, $\mathcal{C}_2 = \mathcal{C}_{0,\infty} \cup \mathcal{C}_{\infty,0}$, and we have the following result for the wealth achieved by the BSP.

THEOREM 3.5 *If Conditions* (B.1) *and* (B.2) *in appendix A holds, then, as* $n \to \infty$,

$$\mathrm{P}\left\{\log \tilde{W}_{n|\mathcal{G}_n} \geq \frac{\Upsilon_{0,\infty}}{\Upsilon_{\infty,\infty} + \Upsilon_{0,\infty}} \log W_{n|\mathcal{G}_n}^{(0)} + \frac{\Upsilon_{\infty,\infty}}{\Upsilon_{\infty,\infty} + \Upsilon_{0,\infty}} \log W_{n|\mathcal{G}_n}^* - u(\Upsilon_{\infty,\infty} + \Upsilon_{\infty,0})\right\} \to 1,$$

where $\mathcal{G}_k = \sigma(\eta_1^*, \eta_2^*, \ldots, \eta_k^*)$ *and* $\Upsilon_{i,j} = \sum \mathrm{E}[|\log \boldsymbol{x}_k^T \boldsymbol{b}_k^{(\infty)} - \log \boldsymbol{x}_k^T \boldsymbol{b}_k^{(0)}| |\mathcal{F}_{k-1}, \mathcal{G}_k]$, *for* $i = 0, \infty$, $j = 0, \infty$, *and the summation over* $k \in \mathcal{C}_{i,j}$. *Moreover, if both Conditions* (B.1) *and* (B.2) *hold, then* $\mathrm{P}\left\{\tilde{W}_{n|\mathcal{G}_n} \geq W_{n|\mathcal{G}_n}^{(0)}\right\} \to 1$ *as* $n \to \infty$.

Remark 10 Condition (B.1) says the classification error rate is less than u, while Condition (B.2) says the bound u should not be too large. Thus, Conditions (B.1) and (B.2) provide an upper bound for the classification error rate of the BSP strategy. If the classification error is 0, then (B.1) and (B.2) hold automatically. The classification error could be interpreted as a risk measure and Conditions (B.1) and (B.2) indicate the risk should be carefully controlled; see also below on measure of risk. It is certainly desirable not to impose structural assumptions on the returns.

Remark 11 Theorem 3.5 shows that, when the classification error is small, the $\mathcal{G}_n-$compressed cumulative wealth using the BSP is asymptotically larger than the weighted average of the compressed wealth of the oracle aggregating algorithm, i.e. the AA using the oracle learning rate, and the one using the 1/N portfolio. Under additional assumptions, the BSP is uniformly better than the 1/N in terms of compressed cumulative wealth.

To this end, we recommend using the K-nearest neighbour method, which is model free. Specifically, for a given K, let

$$\kappa(n) = \begin{cases} 1, & \text{if } n > K+1, \text{ and } \frac{\#\left\{\{n-K,n-K+1,\ldots,n-1\}\cap\mathcal{C}_1\right\}}{\#\{n-K,n-K+1,\ldots,n-1\}} > 0.5, \\ 2, & \text{else.} \end{cases}$$

The description of the BSP algorithm is given as follows:

Algorithm 1 For a given threshold $u > 0, \eta_1 = 0, \eta_2 = \infty$, classification rule κ that satisfies (12), $n \geq 3, \boldsymbol{x}_i, i = 1, 2, \ldots, n-1$,

Step 1 Calculate $\boldsymbol{b}_i^{(\infty)}$ according to (3), $i = 1, 2, \ldots, n$;
Step 2 Calculate η_i^* according to (10), $i = 1, 2, \ldots, n-1$, and $\mathcal{C}_j \cap \{2, 3, \ldots, n-1\}$ according to (11), $j = 1, 2$;
Step 3 Calculate $\kappa(i), i = 2, 3, \ldots, n$ and the empirical classification error $\mathcal{E}_{n-1}(\kappa)$ according to (13).
Step 4 We have the output portfolio at t_n,

$$\boldsymbol{b}_n^{(\tilde{\eta}_n)} = \begin{cases} \boldsymbol{b}_n^{(\eta_{n-1}^*)}, & \text{if } \kappa(n) = 1 \text{ and } \mathcal{E}_{n-1}(\kappa) \leq u; \\ \boldsymbol{b}^{(0)}, & \text{otherwise.} \end{cases}$$

4. Empirical studies

In this section, we show the empirical performance of the BSP for three data-sets, and compare it with those of the 1/N, the CUP, the SCRP and other portfolio construction methods. To evaluate the overall performance of all methods, we calculate and compare the ending cumulative wealth and Sharpe ratio using all methods to select portfolios from each combination of two stocks. We use $(i, j) \in \{(i, j)|i \neq j, i, j = 1, 2, \ldots, p\}$, to represent the combination of the ith and jth stocks. Note that there are $p(p-1)/2$ different combinations. The corresponding returns at t_n could be denoted by $(x_{n,i}, x_{n,j})^T$ for the (i, j) combination. For each method, calculate $\boldsymbol{b}_n^{\langle i,j\rangle}$ at t_n as the portfolio for $(x_{n,i}, x_{n,j})^T$, where $\boldsymbol{b}_n^{\langle i,j\rangle} \in \{\boldsymbol{b} \in \boldsymbol{B}|b_{n,k} = 0, \forall k \notin \{i, j\}\}$. The corresponding cumulative wealth and the empirical Sharpe ratio using combination (i, j) are denoted by

$$W_n^{\langle i,j\rangle} = \prod_{i=1}^{n} \boldsymbol{x}_n^T \boldsymbol{b}_n^{\langle i,j\rangle},$$

$$S^{\langle i,j\rangle} = \frac{\mathrm{E}_n\left[\left(\boldsymbol{x}_1^T \boldsymbol{b}_1^{\langle i,j\rangle}, \ldots, \boldsymbol{x}_n^T \boldsymbol{b}_n^{\langle i,j\rangle}\right)\right] - 1}{\sqrt{\mathrm{Var}_n\left[\left(\boldsymbol{x}_1^T \boldsymbol{b}_1^{\langle i,j\rangle}, \ldots, \boldsymbol{x}_n^T \boldsymbol{b}_n^{\langle i,j\rangle}\right)\right]}}, \quad (14)$$

where E_n, and Var_n indicate the sample mean and variance.

4.1. Hang seng index data-set

The first data-set includes the daily returns of the 59 constituent stocks of the Hong Kong Hang Seng Index (HSI), from 13 March 2008 to 16 April 2013. For each method, a boxplot or a quantile-probability plot of $\{\log W_n^{\langle i,j\rangle}, i, j = 1, \ldots, p, i < j\}$ or $\{S^{\langle i,j\rangle}, i, j = 1, \ldots, p, i < j\}$ is used to characterize the overall performance for each method. Figure 1(a) and (b) show boxplots that compare the cumulative log-wealth and Sharpe ratio defined by (14), using four different methods i.e. the AA with $\eta = 0$, $\eta = 1$, $\eta = \infty$ and the BSP.

We find that, in general, the performance of the AA with $\eta = 0$ (1/N portfolio) and the BSP are better than that of the AA with $\eta = 1$ or ∞. To further compare AA with $\eta = 0$ and the BSP, the quantile-probability plots are shown in figure 2(a) and (b). We find that, for most of $\alpha \in [0, 1]$, the corresponding α quantile of log-wealth and Sharpe ratio of the BSP are above those of the 1/N portfolio; see the exact quantile value and the mean value in table 1, where 'Log-W' and 'SR' represent the log-wealth and empirical Sharpe ratio, respectively.

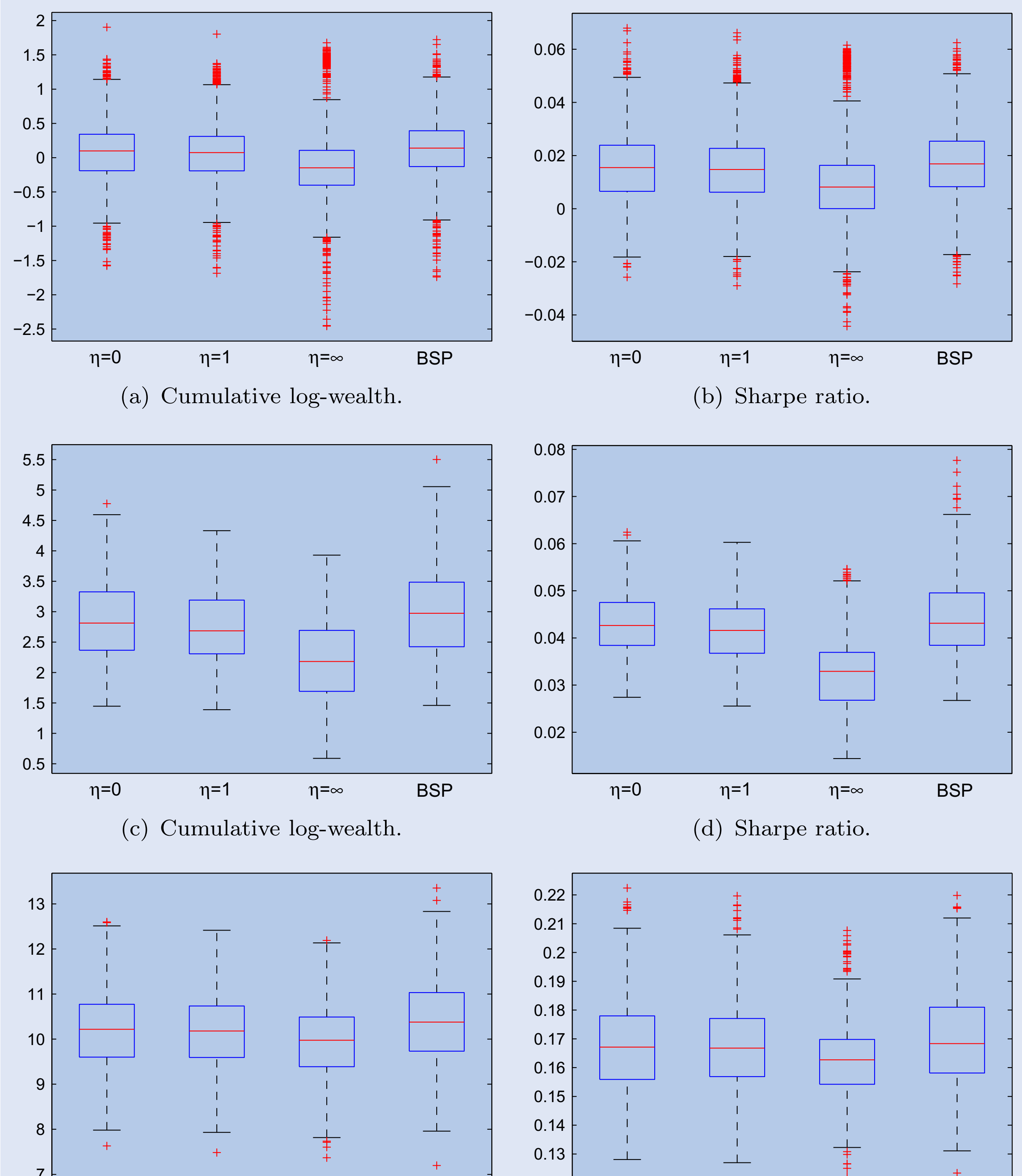
We also report the median of the annual Sharpe ratios in table 2 for the AA with $\eta = 0, 0.05, 0.1, 0.5, 1, \infty$ and the BSP. Here, results using the 1/N method are considered to be the baseline. We observe that the BSP has the largest annual Sharpe ratio for virtually every year. The results of the AA with other learning rates are omitted as they all underperform the BSP.

Lastly, we study the investment on multiple stocks where the BSP is compared with the buy-the-winner strategy (Jegadeesh and Titman 1993). Following Jegadeesh and Titman (1993), we employ the 'J-month/$K-$month' trading strategy, denoted by '(J, K)', and buy the 6 stocks with largest average returns (top decile). At the beginning of each month we invest $1/K$ of current wealth on the six winner stocks. From table 3, we see that among different trading strategies, buy-the-winner with 3-month/3-month strategy has the largest average returns for the first four years, but performs poorly in 2012–2013. On the other hand, the BSP performs competitively and is stable throughout all the periods considered. In this sense, the BSP can be a viable investment strategy.

4.2. New York Stock Exchange data-set

The second data-set contains the daily returns of 36 stocks in the New York Stock Exchange for the period of 1962–1984. This data-set was used by Cover (1991), Helmbold *et al.* (1998) and Gaivoronski and Stella (2000). We compare the resulting using four portfolios, i.e. the AA with $\eta = 0$, 1 and ∞ and the BSP. The corresponding boxplot and the quantile-probability plot are rendered in figure 1(c), (d), 2(c) and (d), with values given in table 4. The annual Sharpe ratio values are reported in table 5. From these results, we arrive at the same conclusion as in the previous subsection that the overall performance of the BSP is the best among the four portfolios in terms of both Sharpe ratio and the cumulative log-wealth.

To gain insights, we now use three representative stock pairs to show the working mechanism of the BSP. The first pair Ford and Schlum tends to be a 'follow-the-leader' type, since more of its η_n^* are ∞. The observation can be verified by examining figure 3, which displays the proportion of Ford in the BSP over time. It is also clear that the cumulative wealth of BSP is much larger than that of others. The pair Fischer and Morrison tends to be a 'half-1/N-half-follow-the-leader' type. Figure 4(b) displays the proportion of Fischer over time using the BSP, which demonstrates that there exist blocks of



(a) Cumulative log-wealth. (b) Sharpe ratio.

(c) Cumulative log-wealth. (d) Sharpe ratio.

(e) Cumulative log-wealth. (f) Sharpe ratio.

Figure 1. Boxplot comparison of the AA with $\eta = 0, 1, \infty$ and the BSP in terms of cumulative log-wealth and Sharpe ratio. Panels (a) and (b): Hong Kong HSI; Panels (c) and (d): NYSE; Panels (e) and (f): Fama–French 30 Industry Classification.

time for the same type of portfolios, including the 1/N. See figure 5 for stocks Iroquois and Kinark. Figure 5(b) shows the proportion of Iroquois over time using the BSP, which is a constant 0.5. The BSP coincides with the 1/N in figure 5(a). Thus, we can conclude, at least for this data-set, the BSP is indeed an adaptive portfolio selection method that can pick the 'optimal' portfolio in all three very different scenarios.

4.3. Fama–French 30 Industry Classification data-set

The third data-set is the monthly return data-set of Fama–French 30 Industry Classifications (Fama and French 1997), which is available from Kenneth French's webpage. The time period is from July 1926 to December 2011. From tables 6 and 7 and figures 1(e), (f), 2(e) and (f), we again observe

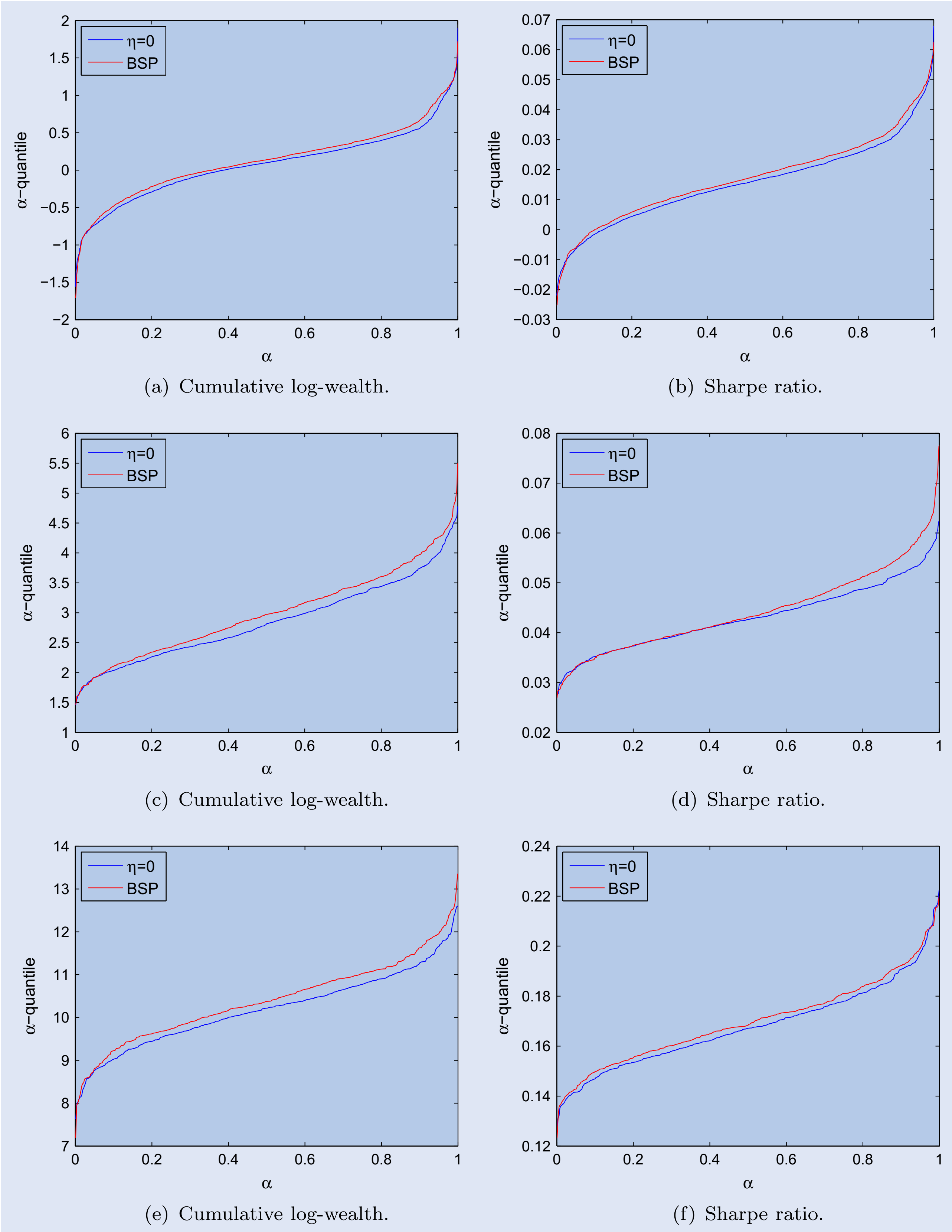


Figure 2. Quantile plot comparison of the AA with $\eta = 0$ and the BSP in terms of cumulative log-wealth and Sharpe ratio. Panels (a) and (b): Hong Kong HSI; Panels (c) and (d): NYSE; Panels (e) and (f): Fama–French 30 Industry Classification.

Table 1. Comparison of the logarithm of cumulative wealth and empirical Sharpe ratio using Hong Kong HSI.

Log-W	Min	.25 quantile	Median	Mean	.75 quantile	Max
$\eta = 0$	−1.5774	−0.1914	0.0985	0.0709	0.3418	1.9029
$\eta = 1$	−1.6867	−0.1922	0.0745	0.0591	0.3115	1.8022
$\eta = \infty$	−2.4567	−0.4017	−0.1480	−0.1148	0.1066	1.6769
BSP	−1.7405	−0.1295	0.1392	0.1240	0.3934	1.7222
SR						
$\eta = 0$	−0.0258	0.0065	0.0155	0.0156	0.0239	0.0679
$\eta = 1$	−0.0290	0.0062	0.0148	0.0149	0.0227	0.0662
$\eta = \infty$	−0.0443	0.0000	0.0082	0.0094	0.0163	0.0615
BSP	−0.0283	0.0083	0.0168	0.0172	0.0254	0.0625

Table 2. Comparison of the increment over the 1/N portfolio on the median of the annual Sharpe ratios using Hong Kong HSI data-set.

Year	$\eta = 0$	$\eta = 0.05$	$\eta = 0.1$	$\eta = 0.5$	$\eta = 1$	$\eta = \infty$	BSP
2008–2009	0	−0.0010	−0.0021	−0.0180	−0.0327	−0.1800	0.0549
2009–2010	0	−0.0005	−0.0011	−0.0140	−0.0299	−0.3609	−0.0109
2010–2011	0	0.0002	0.0031	0.0072	0.0081	0.0528	−0.0075
2011–2012	0	−0.0006	−0.0005	−0.0058	−0.0122	−0.0890	0.0048
2012–2013	0	0.0011	0.0051	0.0147	0.0250	0.0702	0.0296

Table 3. Comparison of the increment over the 1/N portfolio on the median of the annual returns and annual Sharpe ratios using the BSP and other approaches with Hong Kong HSI data-set.

Ann. return	WSCRP(.9995)	SP(.45)	EG(.05)	(3,3)	(6,3)	(6,6)	(12,6)	BSP
2008–2009	−0.0005	0.0068	−0.0022	0.0741	0.1996	0.2108	−0.1216	0
2009–2010	0.1327	−0.0034	−0.0015	0.0480	0.0914	0.0589	−0.2592	0.0129
2010–2011	−0.1461	−0.0005	−0.0006	0.0982	0.2082	0.1874	0.1590	0.0011
2011–2012	0.0284	0.0019	0.0003	0.1026	−0.0428	−0.0443	−0.0585	0.1158
2012–2013	0.0527	0.0006	0.0017	−0.1802	−0.0966	−0.0056	0.0733	0.0513
SR								
2008–2009	−0.0063	0.0170	−0.0093	0.0704	0.2830	0.2033	−0.3803	0
2009–2010	0.2590	−0.0079	−0.0014	−0.3296	−0.2423	−0.2146	−0.6285	0.0264
2010–2011	−0.6959	−0.0020	−0.0032	0.2304	0.5367	0.5030	0.3459	−0.0528
2011–2012	0.1257	0.0064	0.0010	0.3596	−0.0525	−0.0194	−0.0021	0.3999
2012–2013	0.2285	0.0032	0.0096	−0.8590	−0.4734	−0.0484	0.3139	0.2286

Table 4. Comparison of the logarithm of cumulative wealth and empirical Sharpe ratio using NYSE data-set.

Log-W	Min	.25 quantile	Median	Mean	.75 quantile	Max
$\eta = 0$	1.4459	2.3658	2.8133	2.8535	3.3255	4.7765
$\eta = 1$	1.3874	2.3071	2.6855	2.7354	3.1901	4.3311
$\eta = \infty$	0.5884	1.6902	2.1819	2.2289	2.6914	3.9290
BSP	1.4595	2.4238	2.9731	2.9943	3.4853	5.5021
SR						
$\eta = 0$	0.0274	0.0384	0.0427	0.0431	0.0475	0.0624
$\eta = 1$	0.0256	0.0367	0.0416	0.0416	0.0462	0.0603
$\eta = \infty$	0.0144	0.0268	0.0329	0.0326	0.0369	0.0546
BSP	0.0267	0.0384	0.0431	0.0443	0.0496	0.0777

Table 5. Comparison of the increment over the 1/N portfolio on the median of the annual Sharpe ratios using Cover's 62–84 data-set.

Year	$\eta = 0$	$\eta = .05$	$\eta = .1$	$\eta = .5$	$\eta = 1$	$\eta = \infty$	BSP
1962–1963	0	−0.0003	−0.0007	−0.0074	−0.0155	−0.4656	−0.0083
1963–1964	0	−0.0004	−0.0008	−0.0026	−0.0026	−0.1078	−0.0085
1964–1965	0	−0.0018	−0.0035	−0.0151	−0.0285	−0.3493	0.0122
1965–1966	0	−0.0014	−0.0028	−0.0101	−0.0205	−0.2943	0.0148
1966–1967	0	−0.0004	−0.0023	−0.0063	−0.0310	−0.4214	0.0058
1967–1968	0	0.0002	−0.0012	−0.0211	−0.0310	−0.2536	0.0069
1968–1969	0	0.0007	−0.001	−0.0228	−0.0471	−0.1754	0.0264
1969–1970	0	−0.0018	−0.0037	−0.0073	−0.0183	−0.0128	−0.0067
1970–1971	0	0.0022	0.0002	−0.0109	−0.0197	−0.3461	−0.0039
1971–1972	0	0.0202	0.0349	0.0143	0.1521	0.3254	0.1333
1972–1973	0	0.0036	0.0078	0.0068	0.0108	−0.0860	0.0271
1973–1974	0	−0.001	−0.0019	−0.0088	−0.0295	−0.0650	0.0392
1974–1975	0	−0.0037	−0.0082	−0.0262	−0.0477	−0.2384	0.0661
1975–1976	0	−0.001	−0.003	−0.0169	−0.0484	−0.2881	0.0354
1976–1977	0	−0.0039	−0.0028	−0.0336	−0.0988	−0.1485	0.0919
1977–1978	0	0.0004	−0.0001	0.0075	0.0222	0.0706	0.0212
1978–1979	0	−0.0039	−0.0062	−0.0105	−0.0234	−0.0331	−0.0043
1979–1980	0	−0.002	−0.003	0.0161	0.0222	−0.0017	−0.0001
1980–1981	0	−0.0025	−0.0057	−0.0179	−0.0116	−0.0939	0.0362
1981–1982	0	−0.0055	−0.0107	−0.024	−0.0505	−0.1367	0.0341
1982–1983	0	−0.0005	0.0012	−0.019	−0.0342	−0.2701	0.0093
1983–1984	0	−0.0013	−0.0027	−0.0105	−0.0275	−0.0886	0.0083

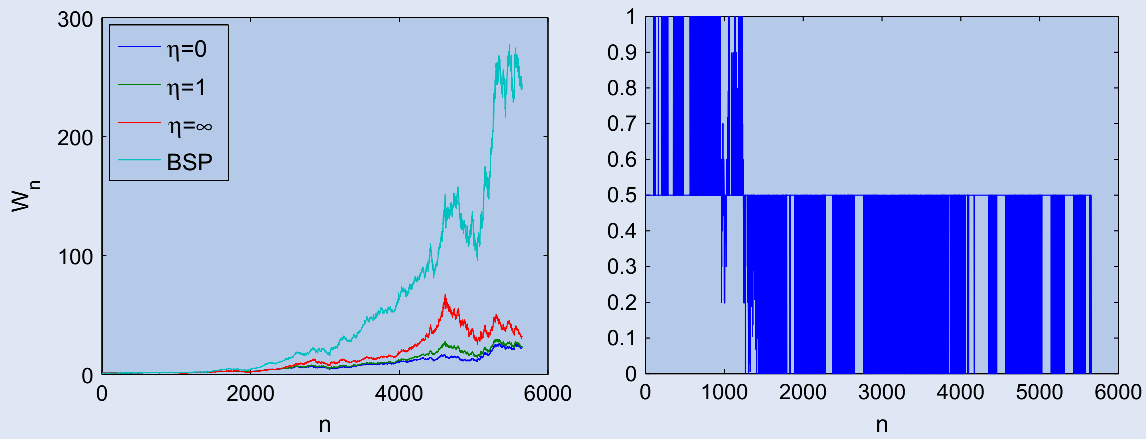
Table 6. Comparison of the logarithm of cumulative wealth and empirical Sharpe ratio using FF30 data-set.

Log-W	Min	.25 quantile	Median	Mean	.75 quantile	Max
$\eta = 0$	7.6302	9.6011	10.2181	10.1843	10.7737	12.6007
$\eta = 1$	7.4838	9.5927	10.1805	10.1581	10.7351	12.4157
$\eta = \infty$	6.7220	9.3893	9.9754	9.9103	10.4905	12.1894
BSP	7.1960	9.7343	10.3770	10.3957	11.0318	13.3515
SR						
$\eta = 0$	0.1281	0.1559	0.1671	0.1678	0.1780	0.2224
$\eta = 1$	0.1270	0.1569	0.1668	0.1677	0.1771	0.2197
$\eta = \infty$	0.1186	0.1542	0.1627	0.1630	0.1698	0.2076
BSP	0.1234	0.1581	0.1683	0.1699	0.1810	0.2198

Table 7. Comparison of the increment over the 1/N portfolio on the median of the annual Sharpe ratios using FF30 monthly return data-set.

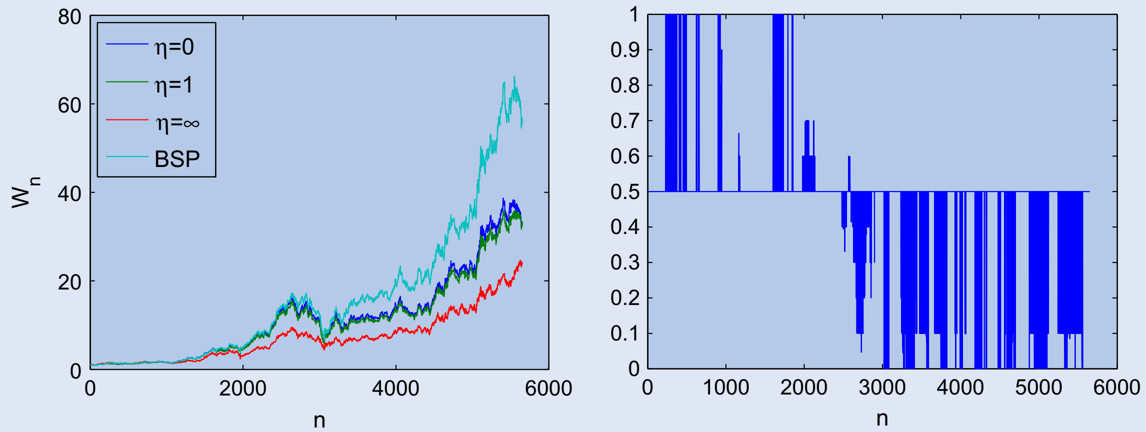
Year	$\eta = 0$	$\eta = 0.05$	$\eta = 0.1$	$\eta = 0.5$	$\eta = 1$	$\eta = \infty$	BSP
1926–1936	0	−0.0001	−0.0004	−0.0018	−0.0038	−0.0237	−0.0025
1936–1946	0	−0.0002	−0.0003	−0.0025	−0.0075	−0.0394	0.0008
1946–1956	0	0.0004	−0.0003	−0.0023	−0.0030	0.0057	0.0188
1956–1966	0	0.0001	−0.0014	−0.0129	−0.0183	−0.1022	−0.0003
1966–1976	0	0.0005	0.0005	0.0006	−0.0003	0.0080	0.0101
1976–1986	0	0.0001	−0.0003	0.0028	−0.0020	−0.0672	0.0016
1986–1996	0	0.0009	0.0018	0.0080	0.0096	0.0412	0.0158
1996–2006	0	0.0001	−0.0004	−0.0038	−0.0039	−0.0241	0.0109
2006–2011	0	−0.0002	0.0010	0.0043	0.0060	0.0343	0.0173

similar results that the BSP outperforms the AA with a fixed learning rate η for a range of different η values. Using this data-set, we also compare the BSP with other methods, i.e. the weighted successive constant rebalanced portfolios (WSCRP) of Gaivoronski and Stella (2000), the switching portfolios (SP) of Singer (1997) and the exponential gradient method (EG) of Helmbold *et al.* (1998). Note that the three methods contain a tuning parameter γ. We see clearly from table 8 that the BSP attains the largest annual Sharpe ratios in most cases.



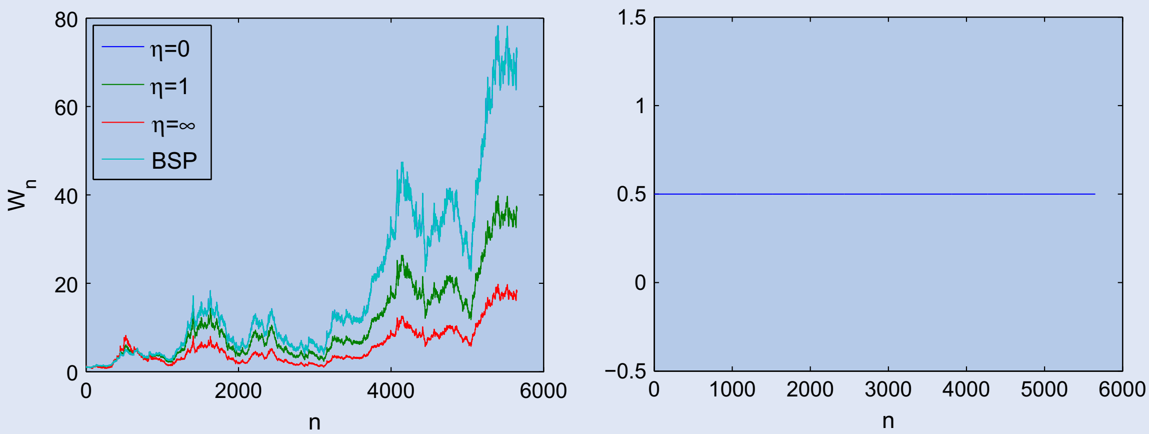
(a) Cumulative wealth (b) Proportion of Ford in the BSP over time

Figure 3. The stock pair Ford and Schlum.



(a) Cumulative wealth (b) Proportion of Fisch in the BSP over time

Figure 4. The stock pair Fisch and Morris.



(a) Cumulative wealth (b) Proportion of Iroqu in the BSP over time

Figure 5. The stock pair Iroqu and Kinark.

Table 8. Comparison of the increment over the 1/N portfolio on the median of the annual Sharpe ratios using FF30.

Year						
WSCRP	$\gamma = .999$	$\gamma = .9995$	$\gamma = .9999$	$\gamma = .99995$	$\gamma = .99999$	BSP
1926–1936	0.0013	0.0009	$-2.43e{-}05$	$1.69e{-}04$	−3e−05	−0.0025
1936–1946	−0.0043	−0.003	$-2.14e{-}04$	$-1.47e{-}04$	$-5.17e{-}06$	0.0008
1946–1956	0.0119	0.013	$2.56e{-}03$	$9.52e{-}04$	$-2.51e{-}04$	0.0188
1956–1966	−0.0182	−0.0066	$-1.71e{-}03$	$-6.94e{-}04$	$-4.46e{-}04$	−0.0003
1966–1976	0.0076	0.0023	$2.88e{-}04$	$-3e{-}04$	$-6.22e{-}05$	0.0101
1976–1986	−0.0285	−0.013	$-2.85e{-}03$	$-9.38e{-}04$	$-3.59e{-}04$	0.0016
1986–1996	0.0154	0.009	$2.17e{-}03$	$9.5e{-}04$	$1.76e{-}04$	0.0158
1996–2006	−0.032	−0.013	$-5.75e{-}04$	$-7.34e{-}05$	$3.89e{-}04$	0.0109
2006–2011	0.001	0.0034	$-8.56e{-}04$	$-3.66e{-}04$	$5.24e{-}04$	0.0173
SP	$\gamma = .025$	$\gamma = .045$	$\gamma = .06$	$\gamma = .075$	$\gamma = .1$	
1926–1936	0.00065	0.0024	0.0025	0.0013	0.0008	−0.0025
1936–1946	0.0016	0.0026	0.0021	0.0021	0.0015	0.0008
1946–1956	0.0023	0.0057	0.0055	0.0049	0.0026	0.0188
1956–1966	0.0068	0.0076	0.008	0.0066	0.0045	−0.0003
1966–1976	0.0078	0.0075	0.0067	0.0062	0.0057	0.0101
1976–1986	0.0023	0.0071	0.0044	0.004	0.0044	0.0016
1986–1996	0.0068	0.009	0.0081	0.0067	0.0044	0.0158
1996–2006	−0.0038	0.00074	0.0007	0.0017	0.0021	0.0109
2006–2011	−0.0026	−0.00057	0.001	0.001	0.0034	0.0173
EG	$\gamma = .02$	$\gamma = .035$	$\gamma = .05$	$\gamma = .065$	$\gamma = .08$	
1926–1936	0.00027	0.00047	0.00051	0.00022	−0.00007	−0.0025
1936–1946	−0.00015	−0.00011	−0.00007	−0.00034	−0.00028	0.0008
1946–1956	0.00002	0.00298	0.00314	0.00328	0.00233	0.0188
1956–1966	−0.00124	−0.00229	−0.00328	−0.00426	−0.00550	−0.0003
1966–1976	−0.00037	−0.00052	−0.00066	−0.00189	−0.00188	0.0101
1976–1986	−0.00014	−0.00057	−0.00185	−0.00453	−0.00593	0.0016
1986–1996	−0.00004	−0.00007	−0.00010	0.00088	0.00203	0.0158
1996–2006	−0.00157	−0.00275	−0.00393	−0.00513	−0.00406	0.0109
2006–2011	0.00071	0.00134	0.00197	0.00260	0.00196	0.0173

5. Concluding remarks

This paper proposes a new portfolio construction method that switches between two opposing investment strategies: one momentum-based and the other mean reverting-oriented. It is motivated by the intuition that there exist blocks of periods in which either the momentum-based or the mean reverting approach is preferred and that the blocks may be identified via suitable machine learning algorithms. Theoretical justification as well as empirical evidence show the superiority of the new method.

In practice, there may exist related macro- and micro-variables. It would be interesting to see how such variables can be incorporated into the learning algorithms. Furthermore, our analysis here does not take into account transaction costs, which could make the situation considerably more complicated.

Acknowledgements

The authors thank the Editor, the AE and two anonymous referees for their insightful comments which have greatly improved the paper. The research of Tengfei Li and Zhiliang Ying was partially supported by Shanghai Center for Mathematical Sciences at Fudan University.

Disclosure statement

The authors report no declarations of interest.

Funding

The research of Kani Chen was supported by Hong Kong RGC [grant number 601011], [grant number 600612], [grant number 600813], [grant number 163714]. The research of Yang Feng was partially supported by NSF grant DMS-1308566, NSF CAREER [grant number DMS-1554804]; Lenfest Junior Faculty Development Grant from Columbia University.

References

Belentepe, C.Y. and Wyner, A.J., A statistical view of universal stock market portfolios. Proceedings of the International Symposium on Information Theory, Adelaide, Australia, 573–577, 2005.

Cover, T.M., Universal portfolios. *Math. Finance*, 1991, **1**, 1–29.

Cover, T.M., Universal data compression and portfolio selection. *Proc. FOCS*, 1996, **96**, 534.

Cover, T.M. and Ordentlich, E., Universal portfolios with side information. *IEEE Trans. Inf. Theory*, 1996, **42**, 348–363.

Cross, J.E. and Barron, A.R., Efficient universal portfolios for past-dependent target classes. *Math. Finance*, 2003, **13**, 245–276.

DeMiguel V., Garlappi L. and Uppal R., Optimal vs. naive diversification: How inefficient is the 1/N portfolio strategy? *Rev. Financial Stud.*, 2009, **22**, 1915–1953.

Fama, E.F. and French, K.R., Industry costs of equity. *J. Financ. Econ.*, 1997, **43**, 153–193.

Gaivoronski, A.A. and Stella, F., Stochastic nonstationary optimization for finding universal portfolios. *Ann. Oper. Res.*, 2000, **100**, 165–188.

Helmbold, D.P., Schapire, R.E., Singer, Y. and Warmuth, M.K., On-line portfolio selection using multiplicative updates. *Math. Finance*, 1998, **8**, 325–347.

Huang, W., Nakamori, Y. and Wang, S.Y., Forecasting stock market movement direction with support vector machine. *Comput. Oper. Res.*, 2005, **32**, 2513–2522.

Hwang, S. and Rubesam, A., The disappearance of momentum. *Eur. J. Finance*, 2015, 21, 584–607.

Jamshidian, F., Asymptotically optimal portfolios. *Math. Finance*, 1992, **2**, 131–150.

Jegadeesh, N. and Titman, S., Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance*, 1993, **48**, 65–91.

Kalai, A. and Vempala, S., Efficient algorithms for universal portfolios. *J. Mach. Learn. Res.*, 2003, **3**, 423–440.

Kelly, J.L. Jr, A new interpretation of information rate. *IRE Trans. Inf. Theory*, 1956, **2**, 185–189.

Littlestone N. and Warmuth M.K., The weighted majority algorithm. 30th Annual Symposium on Proceedings of the Foundations of Computer Science, Research Triangle Park, NC256–261, 1989.

Markowitz, H., Portfolio selection. *J. Finance*, 1952, **7**, 77–91.

Ordentlich E. and Cover T.M., On-line portfolio selection. Proceedings of the Ninth Annual Conference on Computational Learning Theory, New York, NY, 310–313, 1996.

Patel, J., Shah, S., Thakkar, P. and Kotecha, K., Predicting stock market index using fusion of machine learning techniques. *Expert Syst. Appl.*, 2015, **42**, 2162–2172.

Pflug, G.C., Pichler, A. and Wozabal, D., The 1/N investment strategy is optimal under high model ambiguity. *J. Bank. Finance*, 2012, **36**, 410–417.

Qian, B. and Rasheed, K., Stock market prediction with multiple classifiers. *Appl. Intell.*, 2006, **26**, 25–33.

Singer, Y., Switching portfolios. *Int. J. Neural Syst.*, 1997, **8**, 445–455.

Tu, J. and Zhou, G., Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *J. Financ. Econ.*, 2011, **99**, 204–215.

Vovk V.G., Aggregating strategies. Proceedings of the Third Workshop on Computational Learning Theory, Rochester, NY, 371–383, 1990.

Vovk V. and Watkins C., Universal portfolio selection. Proceedings of the Eleventh Annual Conference on Computational Learning Theory, Madison, WI, 12–23, 1998.

Appendix 1. Conditions

We introduce several additional conditions that are used to prove the theoretical results.

(A.4) All stocks are strictly active: there exist positive constants N and c_3, such that for any $n > N, b^*_{n,j} \ge c_3,\ j = 1, 2, \ldots, p$.

(A.5) There exist positive constants N, c_4 and c_5, such that for any $n > N$, $\log W_n(\boldsymbol{b}^*_n)/n \in [c_4, c_5]$.

(A.6) There exist positive constants N, c_6, such that for any $n > N$,

$$\left|\boldsymbol{u}_0'^T \tilde{\mathbf{J}}_1^{-1}\boldsymbol{u}_0' - (p-2)\mathrm{trace}(\tilde{\mathbf{J}}_1^{-1}\tilde{\mathbf{J}}_0) - \gamma_n(p-1)^2/2\right| > c_6$$

where

$$\tilde{\mathbf{J}}_0 = \frac{(\boldsymbol{x}'_{n+1} - x_{n+1,p}\mathbf{1}_p)(\boldsymbol{x}'_{n+1} - x_{n+1,p}\mathbf{1}_p)^T}{(\boldsymbol{x}^T_{n+1}\boldsymbol{b}^*_n)^2};$$

$$\tilde{\mathbf{J}}_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{(\boldsymbol{x}'_i - x_{i,p}\mathbf{1}_p)(\boldsymbol{x}'_i - x_{i,p}\mathbf{1}_p)^T}{(\boldsymbol{x}^T_i\boldsymbol{b}^*_n)^2},$$

$$\boldsymbol{u}'_0 = \left(u'_{0,1}, \ldots, u'_{0,p-1}\right),\ \gamma_n = \frac{n}{\log W_n(\boldsymbol{b}^*_n)},$$

where $\boldsymbol{x}'_i = (x_{i,1}, \ldots, x_{i,p-1})^T$ and $u'_{0,j} = (x_{n+1,j} - x_{n+1,p})/(\boldsymbol{x}^T_{n+1}\boldsymbol{b}^*_{n+1}),\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, p-1$.

(A.7) There exist positive constants N and c_7, such that, for any $n > N$,

$$\frac{1}{n}\left|\int_{\boldsymbol{B}}\boldsymbol{x}^T_n\boldsymbol{b}\log W_{n-1}(\boldsymbol{b})\mathrm{d}\pi(\boldsymbol{b}) - \int_{\boldsymbol{B}}\log W_{n-1}(\boldsymbol{b})\mathrm{d}\pi(\boldsymbol{b}) \times \int_{\boldsymbol{B}}\boldsymbol{x}^T_n\boldsymbol{b}\mathrm{d}\pi(\boldsymbol{b})\right| > c_7.$$

(B.1) $\mathcal{E}(\kappa) \le u$, where

$$\mathcal{E}(\kappa) = \overline{\lim}\, \mathrm{P}\big(n \in \mathcal{C}_1, \kappa(n) = 2 \text{ or } n \in \mathcal{C}_2, \kappa(n) = 1\big)$$

represents the upper limit of the classification error as $n \to \infty$.

(B.2)

$$\lim_{n\to\infty}\mathrm{P}\left\{\frac{\Upsilon_{\infty,\infty}}{\Upsilon_{\infty,\infty} + \Upsilon_{\infty,0}} \ge u\right\} = 1.$$

Appendix 2. Proofs

Proof of Lemma 2.1 For any $\delta_0 > 0$, denote

$$B_{\delta_0}(\boldsymbol{b}^*) = \{\boldsymbol{b} \in B \big| \|\boldsymbol{b} - \boldsymbol{b}^*\| < \delta_0\}.$$

Let

$$\tilde{\boldsymbol{b}}_n(\delta_0) = \arg\max_{\boldsymbol{b}\in B\backslash B_{\delta_0}(\boldsymbol{b}^*_n)} W_n(\boldsymbol{b}) \text{ and } \zeta_n = \sqrt{\frac{2p}{(1-2\epsilon)\eta\delta}\frac{\log n}{n}}, \tag{B1}$$

where ϵ is an arbitrary constant in (0, 0.5). Note that $B\backslash B_{\delta_0}(\boldsymbol{b}^*_n)$ is the compact complement of $B_{\delta_0}(\boldsymbol{b}^*_n)$ within B. From theorem 2 and equation (16) in Gaivoronski and Stella (2000), we have

$$\frac{1}{n}\log W_n\big(\tilde{\boldsymbol{b}}_n(\zeta_n)\big) - \frac{1}{n}\log W_n(\boldsymbol{b}^*_n) \le \frac{\partial\left[\frac{1}{n}\log W_n(\boldsymbol{b}^*_n)\right]}{\partial\boldsymbol{b}^T}(\tilde{\boldsymbol{b}}_n(\zeta_n) - \boldsymbol{b}^*_n) - \frac{\delta\|\tilde{\boldsymbol{b}}_n(\zeta_n) - \boldsymbol{b}^*_n\|^2}{2}. \tag{B2}$$

If $\boldsymbol{b}^*_n$ is an inner point of B, $\frac{\partial[\frac{1}{n}\log W_n(\boldsymbol{b}^*_n)]}{\partial\boldsymbol{b}} = 0$. On the other hand, if $\boldsymbol{b}^*_n$ is on the boundary of B,

$$\frac{\partial\left[\frac{1}{n}\log W_n(\boldsymbol{b}^*_n)\right]}{\partial\boldsymbol{b}^T}(\tilde{\boldsymbol{b}}_n(\zeta_n) - \boldsymbol{b}^*_n) = \frac{\partial\left[\frac{1}{n}\log W_n\Big(\boldsymbol{b}^*_n + \beta\big(\tilde{\boldsymbol{b}}_n(\zeta_n) - \boldsymbol{b}^*_n\big)\Big)\right]}{\partial\beta} = \lim_{\beta\to 0+}\frac{\frac{1}{n}\log W_n\Big(\boldsymbol{b}^*_n + \beta\big(\tilde{\boldsymbol{b}}_n(\zeta_n) - \boldsymbol{b}^*_n\big)\Big) - \frac{1}{n}\log W_n\big(\boldsymbol{b}^*_n\big)}{\beta} \le 0,$$

due to the optimality of $\boldsymbol{b}^*_n$. Continue with (B2), we have

$$\log W_n\big(\tilde{\boldsymbol{b}}_n(\zeta_n)\big) - \log W_n(\boldsymbol{b}^*_n) \le -\frac{n\delta\|\tilde{\boldsymbol{b}}_n(\zeta_n) - \boldsymbol{b}^*_n\|^2}{2} \le -\frac{n\delta\zeta_n^2}{2}. \tag{B3}$$

Choose $\varsigma_n = (\delta\zeta_n^2 c_1\epsilon)/(2c_2\sqrt{p})$. Then, $\varsigma_n \to 0$ as $n \to \infty$ by the definition of ζ_n in (B1). For any $\boldsymbol{b} \in B_{\varsigma_n}(\boldsymbol{b}^*_n)$, it follows from the mean value theorem and Cauchy inequality that there exists some $\boldsymbol{b}_1 \in B$, such that

$$\log W_n\big(\boldsymbol{b}\big) - \log W_n(\boldsymbol{b}^*_n) = \frac{\partial\log W_n(\boldsymbol{b}_1)}{\partial\boldsymbol{b}^T}(\boldsymbol{b} - \boldsymbol{b}^*_n) \ge -\varsigma_n\sqrt{\frac{\partial\log W_n(\boldsymbol{b}_1)}{\partial\boldsymbol{b}^T}\frac{\partial\log W_n(\boldsymbol{b}_1)}{\partial\boldsymbol{b}}}.$$

Condition (A.2) ensures $\|\partial\log W_n(\boldsymbol{b})/\partial\boldsymbol{b}\| \le nc_2\sqrt{p}/c_1$. Then

$$\log W_n\big(\boldsymbol{b}\big) - \log W_n(\boldsymbol{b}^*_n) \ge -\frac{nc_2\sqrt{p}}{c_1}\varsigma_n = -\frac{n\delta\zeta_n^2\epsilon}{2}. \tag{B4}$$

From (B3) and (B4), we have $\varsigma_n < \zeta_n$, and for any $\boldsymbol{b} \in B_{\varsigma_n}(\boldsymbol{b}^*_n)$, $\log W_n\big(\boldsymbol{b}\big) - \log W_n(\tilde{\boldsymbol{b}}_n(\zeta_n)) = [\log W_n\big(\boldsymbol{b}\big) - \log W_n\big(\boldsymbol{b}^*_n\big)] + [\log W_n\big(\boldsymbol{b}^*_n\big) - \log W_n(\tilde{\boldsymbol{b}}_n(\zeta_n))] \ge n\delta\zeta_n^2(1-\epsilon)/2$.

By some straightforward calculation,

$$\lim_{n\to\infty}\frac{\int_{B\backslash B_{\zeta_n}(\boldsymbol{b}_n^*)}(b_j)^i[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(b_j)^i[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}$$
$$\le \lim_{n\to\infty}[W_n(\tilde{\boldsymbol{b}}_n(\zeta_n))]^\eta\frac{\int_{B\backslash B_{\zeta_n}(\boldsymbol{b}_n^*)}(b_j)^i\pi(\mathrm{d}\boldsymbol{b})}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(b_j)^i[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}$$
$$=\lim_{n\to\infty}\frac{\int_{B\backslash B_{\zeta_n}(\boldsymbol{b}_n^*)}(b_j)^i\pi(\mathrm{d}\boldsymbol{b})}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(b_j)^i[W_n(\boldsymbol{b})/W_n(\tilde{\boldsymbol{b}}_n(\zeta_n))]^\eta\pi(\mathrm{d}\boldsymbol{b})}$$
$$\le\lim_{n\to\infty}\frac{\int_B(b_j)^i\pi(\mathrm{d}\boldsymbol{b})}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(b_j)^i[W_n(\boldsymbol{b})/W_n(\tilde{\boldsymbol{b}}_n(\zeta_n))]^\eta\pi(\mathrm{d}\boldsymbol{b})}$$
$$\le\frac{\int_B(b_j)^i\pi(\mathrm{d}\boldsymbol{b})}{\lim_{n\to\infty}\exp(\frac{n\delta\zeta_n^2(1-\epsilon)\eta}{2})\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(b_j)^i\pi(\mathrm{d}\boldsymbol{b})}\triangleq\Delta, \quad \text{(B5)}$$

which is related to the power $i=0,1$ and subscript $j=1,2,\ldots,p$.

For the uniform prior $\pi=\pi_0$,

$$\Delta\le\frac{1}{\lim_{n\to\infty}\exp(\frac{n\delta\zeta_n^2\eta}{2}(1-\epsilon))h(\varsigma_n)}=O(n^{-p\frac{1-\epsilon}{1-2\epsilon}}n^p)\to 0 \quad \text{(B6)}$$

holds uniformly in η, where $h(\varsigma_n)=\varsigma_n^p/p^{2p-1}$. To understand why, we assume, without loss of generality, $j=p$. Observe

$$\min_{\boldsymbol{b}^*}\int_{B_{\varsigma n}(\boldsymbol{b}^*)}(b_p)^i\pi_0(\mathrm{d}\boldsymbol{b})\ge\min_{\boldsymbol{b}^*}\int_{B_{\varsigma n}(\boldsymbol{b}^*)}b_p\pi_0(\mathrm{d}\boldsymbol{b}),$$

which could be seen as the volume of an approximate cylinder with the top in the area of B, and with the minimum value achieved at the boundary of B. That is,

$$\min_{\boldsymbol{b}^*}\int_{B_{\varsigma n}(\boldsymbol{b}^*)}b_p\pi_0(\mathrm{d}\boldsymbol{b})\ge\int_{B_{\varsigma n}(\boldsymbol{e}_1)}b_p\pi_0(\mathrm{d}\boldsymbol{b})$$
$$\ge\int_{1-\frac{\varsigma_n}{p}}^{1-\frac{p-2}{p^2}\varsigma_n}\left[1-\sum_{j=1}^{p-1}b_j\right]\mathrm{d}b_1\int_0^{\frac{\varsigma_n}{p^2}}\mathrm{d}b_2\int_0^{\frac{\varsigma_n}{p^2}}\mathrm{d}b_3$$
$$\cdots\int_0^{\frac{\varsigma_n}{p^2}}\mathrm{d}b_{p-1}=\frac{\varsigma_n^p}{p^{2p-1}}.$$

where $\boldsymbol{e}_j$ is a $p\times 1$ vector with the jth element being 1 and all the rest being 0. It is straightforward to show that the above inequality holds for $i=0,1$ and p substituted by $j=1,2,\ldots,p-1$.

For a general continuous prior π, let $f(\cdot)$ be the density function of $\pi(\cdot)$. Since the support of the prior is the compact set B, there exists a constant $C_2>0$ with $\inf_{\boldsymbol{b}\in B}f(\boldsymbol{b})\ge C_2$. Then,

$$\Delta\le\frac{\int_B(b_j)^0\pi(\mathrm{d}\boldsymbol{b})}{\lim_{n\to\infty}\exp(\frac{n\delta\zeta_n^2\eta}{2}(1-\epsilon))\int_{B_{\delta_1}(\boldsymbol{b}_n^*)}(b_j)^i\pi(\mathrm{d}\boldsymbol{b})}$$
$$\le\frac{\int_B(b_j)^0\pi(\mathrm{d}\boldsymbol{b})}{\lim_{n\to\infty}\exp(\frac{n\delta\zeta_n^2\eta}{2}(1-\epsilon))C_2\int_{B_{\delta_1}(\boldsymbol{b}_n^*)}(b_j)^i\pi_0(\mathrm{d}\boldsymbol{b})}.$$

It follows from (B6) that, $\Delta=O(n^{-p\frac{\epsilon}{1-2\epsilon}})\to 0$. Recall the definition of $\boldsymbol{b}_{n+1}^\eta$ in (3), we have

$$\lim_{n\to\infty}\|\boldsymbol{b}_{n+1}^\eta-\boldsymbol{b}_n^*\|$$
$$=\lim_{n\to\infty}\left\|\frac{\int\boldsymbol{b}[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}{\int[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}-\boldsymbol{b}_n^*\right\|$$
$$=\left\|\lim_{n\to\infty}\frac{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\boldsymbol{b}[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})+\int_{B\backslash B_{\zeta_n}(\boldsymbol{b}_n^*)}\boldsymbol{b}[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})+\int_{B\backslash B_{\zeta_n}(\boldsymbol{b}_n^*)}[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}-\boldsymbol{b}_n^*\right\|$$
$$=\left\|\lim_{n\to\infty}\frac{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\boldsymbol{b}[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}-\boldsymbol{b}_n^*\right\|$$
$$=\lim_{n\to\infty}\left\|\frac{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}-\boldsymbol{b}_n^*)[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}\right\|. \quad \text{(B7)}$$

Since $\Delta\to 0$ uniformly in η, (B7) holds uniformly in η. The conclusion then follows from the fact that

$$\left\|\frac{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}-\boldsymbol{b}_n^*)[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}[W_n(\boldsymbol{b})]^\eta\pi(\mathrm{d}\boldsymbol{b})}\right\|\le\zeta_n.$$

□

Proof of Lemma 3.1 The derivative of $\log y_n(\eta)$ is

$$\frac{\partial\log y_n(\eta)}{\partial\eta}=\frac{\partial\log\int\boldsymbol{x}_n^T\boldsymbol{b}W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}{\partial\eta}-\frac{\partial\log\int W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}{\partial\eta}$$
$$=\frac{\int\boldsymbol{x}_n^T\boldsymbol{b}W_{n-1}^\eta(\boldsymbol{b})\log W_{n-1}(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}{\int\boldsymbol{x}_n^T\boldsymbol{b}W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}-\frac{\int W_{n-1}^\eta(\boldsymbol{b})\log W_{n-1}(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}{\int W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}$$
$$=\frac{\int W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}{\int\boldsymbol{x}_n^T\boldsymbol{b}W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}\left[\frac{\int\boldsymbol{x}_n^T\boldsymbol{b}W_{n-1}^\eta(\boldsymbol{b})\log W_{n-1}(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}{\int W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}-\frac{\int W_{n-1}^\eta(\boldsymbol{b})\log W_{n-1}(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}{\int W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}\frac{\int\boldsymbol{x}_n^T\boldsymbol{b}W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}{\int W_{n-1}^\eta(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})}\right]$$
$$=\mathrm{Cov}_\eta\left(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^\eta,\log W_{n-1}(\mathfrak{b}_{n-1}^\eta)\right)/\mathrm{E}_\eta(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^\eta).$$

As a result, the conclusion holds. □

Proof of Theorem 2.3 Define

$$\Delta_n=\log W_n^{(\eta)}-\log W_n(\boldsymbol{b}_n^*).$$

Then,

$$\Delta_{n+1}=\Delta_n+\log x_{n+1}^T\boldsymbol{b}_{n+1}^{(\eta)}-\left[\log W_{n+1}(\boldsymbol{b}_{n+1}^*)-\log W_n(\boldsymbol{b}_n^*)\right]$$
$$=\Delta_n+\left[\log x_{n+1}^T\boldsymbol{b}_{n+1}^{(\eta)}-\log x_{n+1}^T\boldsymbol{b}_{n+1}^*\right]+\log W_n(\boldsymbol{b}_n^*)-\log W_n(\boldsymbol{b}_{n+1}^*)$$
$$\ge\Delta_n+\left[\log x_{n+1}^T\boldsymbol{b}_{n+1}^{(\eta)}-\log x_{n+1}^T\boldsymbol{b}_{n+1}^*\right]$$
$$\ge\Delta_n-\frac{c_2\sqrt{p}}{c_1}\left\|\boldsymbol{b}_{n+1}^{(\eta)}-\boldsymbol{b}_{n+1}^*\right\|.$$

With lemma 2.1 and lemma 1 in Gaivoronski and Stella (2000), one can write

$$\Delta_n\ge\Delta_{n-1}-\frac{c_2\sqrt{p}}{c_1}\|\boldsymbol{b}_n^{(\eta)}-\boldsymbol{b}_n^*\|$$
$$\ge\Delta_{n-1}-\frac{c_2\sqrt{p}}{c_1}\left[\|\boldsymbol{b}_n^{(\eta)}-\boldsymbol{b}_{n-1}^*\|+\|\boldsymbol{b}_{n-1}^*-\boldsymbol{b}_n^*\|\right]$$
$$\ge\Delta_{n-1}-\frac{c_2\sqrt{p}}{c_1}\left[\sqrt{\frac{2p}{\delta\eta}\frac{\log(n-1)}{n-1}}+\frac{2c_2\sqrt{p}}{c_1\delta}\frac{1}{n-1}\right]$$
$$\ge\Delta_1-\frac{c_2p}{c_1}\sqrt{\frac{2}{\delta\eta}}\sum_{k=1}^{n-1}\sqrt{\frac{\log k}{k}}-\frac{2c_2^2p}{c_1^2\delta}\sum_{k=1}^{n-1}\frac{1}{k}.$$

Since the summation of the second term on the right side of the inequality above can be approximated by $\int_0^n(\log t/t)^{1/2}\mathrm{d}t$, and $\int_0^n(\log t/t)^{1/2}\mathrm{d}t=2(n\log n)^{1/2}(1+o(1))$ as $n\to\infty$, it follows that

$$\Delta_n\ge-[\frac{2\sqrt{2}c_2p}{c_1\sqrt{\delta\eta}}\sqrt{n\log n}+\frac{2c_2^2p}{c_1^2\delta}\log n](1+o(1)).$$

Proof of part (b) of the theorem is analogous and is thus omitted. □

Proof of Theorem 3.2 We only need to prove part (b). We use the notation $h(n) \sim g(n)$ to represent that

$$\lim_{n\to\infty} \frac{h(n)}{g(n)} = 1$$

for sequences $h(n)$ and $g(n)$. Using Taylor expansion at $\boldsymbol{b}_n^*$ up to the sixth order, we have

$$\begin{aligned}
\log\left[\boldsymbol{x}_{n+1}^T \boldsymbol{b}\right] &\sim \log\left[\boldsymbol{x}_{n+1}^T \boldsymbol{b}_n^*\right] + \boldsymbol{u}_0^T\left(\boldsymbol{b}-\boldsymbol{b}_n^*\right) \\
&\quad - \frac{1}{2}\left(\boldsymbol{b}-\boldsymbol{b}_n^*\right)^T \mathbf{J}_0 \left(\boldsymbol{b}-\boldsymbol{b}_n^*\right) \\
&\quad + \sum_{j_1,j_2,j_3} w^{(1)}_{j_1,j_2,j_3}\left(b_{j_1}-b^*_{n,j_1}\right)\left(b_{j_2}-b^*_{n,j_2}\right)\left(b_{j_3}-b^*_{n,j_3}\right) \\
&\quad + \cdots \\
&\quad + \sum_{j_1,\ldots,j_6} w^{(1)}_{j_1,\ldots,j_6}\prod_{k=1}^{6}\left(b_{j_k}-b^*_{n,j_k}\right) \triangleq V_1, \\
\log[W_n(\boldsymbol{b})] &\sim \log\left[W_n(\boldsymbol{b}_n^*)\right] - \frac{n}{2}\left(\boldsymbol{b}-\boldsymbol{b}_n^*\right)^T \mathbf{J}_1 \left(\boldsymbol{b}-\boldsymbol{b}_n^*\right) \\
&\quad + \sum_{j_1,j_2,j_3} w^{(2)}_{j_1,j_2,j_3}\left(b_{j_1}-b^*_{n,j_1}\right)\left(b_{j_2}-b^*_{n,j_2}\right)\left(b_{j_3}-b^*_{n,j_3}\right) \\
&\quad + \cdots \\
&\quad + \sum_{j_1,\ldots,j_6} w^{(2)}_{j_1,\ldots,j_6}\prod_{k=1}^{6}\left(b_{j_k}-b^*_{n,j_k}\right) \triangleq V_2, \\
\log\log[W_n(\boldsymbol{b})] &\sim \log\log\left[W_n(\boldsymbol{b}_n^*)\right] - \frac{1}{2}\left(\boldsymbol{b}-\boldsymbol{b}_n^*\right)^T \mathbf{J}_2 \left(\boldsymbol{b}-\boldsymbol{b}_n^*\right) \\
&\quad + \sum_{j_1,j_2,j_3} w^{(3)}_{j_1,j_2,j_3}\left(b_{j_1}-b^*_{n,j_1}\right)\left(b_{j_2}-b^*_{n,j_2}\right)\left(b_{j_3}-b^*_{n,j_3}\right) \\
&\quad + \cdots \\
&\quad + \sum_{j_1,\ldots,j_6} w^{(3)}_{j_1,\ldots,j_6}\prod_{k=1}^{6}\left(b_{j_k}-b^*_{n,j_k}\right) \triangleq V_3,
\end{aligned}$$

where

$$\boldsymbol{u}_0 = \frac{\boldsymbol{x}_{n+1}}{\boldsymbol{x}_{n+1}^T\boldsymbol{b}_n^*}, \mathbf{J}_0 = \frac{\boldsymbol{x}_{n+1}\boldsymbol{x}_{n+1}^T}{\left(\boldsymbol{x}_{n+1}^T\boldsymbol{b}_n^*\right)^2}, \mathbf{J}_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{\boldsymbol{x}_i\boldsymbol{x}_i^T}{\left(\boldsymbol{x}_i^T\boldsymbol{b}_n^*\right)^2},$$

and $\mathbf{J}_2 = \gamma_n^* J_1$ with

$$\gamma_n^* = \frac{n}{\log^2 W_n(\boldsymbol{b}_n^*)} + \frac{n}{\log W_n(\boldsymbol{b}_n^*)} \sim \gamma_n = \frac{n}{\log W_n(\boldsymbol{b}_n^*)}.$$

Notice that γ_n is bounded from 0 and ∞ under Condition (A.5).

The rest of the proof is divided into three parts.

(I) In the first part, we will prove that, given any sequence

$$\zeta_n = \sqrt{\frac{2p}{(1-2\epsilon)\delta\eta_n}\frac{\log n}{n}},\ \eta_n = O(1),\ \frac{(\log n)^\alpha}{n} = o(\eta_n),$$

where $\alpha \in (1, 7]$ is a constant, we have

$$\begin{aligned}
&\frac{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_{n+1})W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b}}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_{n+1})W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b}\times\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}} - 1 \sim U_n \qquad \text{(B8)}\\
&\triangleq \frac{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)} e^{-\frac{n}{2}(\boldsymbol{b}-\boldsymbol{b}_n^*)^T(\frac{\mathbf{J}_0}{n}+\eta_n\mathbf{J}_1+\frac{\mathbf{J}_2}{n})(\boldsymbol{b}-\boldsymbol{b}_n^*)+\boldsymbol{u}_0^T(\boldsymbol{b}-\boldsymbol{b}_n^*)}\mathrm{d}\boldsymbol{b}\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)} e^{-\frac{n}{2}(\boldsymbol{b}-\boldsymbol{b}_n^*)^T\eta_n\mathbf{J}_1(\boldsymbol{b}-\boldsymbol{b}_n^*)}\mathrm{d}\boldsymbol{b}}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)} e^{-\frac{n}{2}(\boldsymbol{b}-\boldsymbol{b}_n^*)^T(\frac{\mathbf{J}_0}{n}+\eta_n\mathbf{J}_1)(\boldsymbol{b}-\boldsymbol{b}_n^*)+\boldsymbol{u}_0^T(\boldsymbol{b}-\boldsymbol{b}_n^*)}\mathrm{d}\boldsymbol{b}\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)} e^{-\frac{n}{2}(\boldsymbol{b}-\boldsymbol{b}_n^*)^T(\eta_n\mathbf{J}_1+\frac{\mathbf{J}_2}{n})(\boldsymbol{b}-\boldsymbol{b}_n^*)}\mathrm{d}\boldsymbol{b}} - 1,
\end{aligned}$$

and

$$U_n = \frac{\gamma_n[\boldsymbol{u}_0'^T\tilde{\mathbf{J}}_1^{-1}\boldsymbol{u}_0' - (p-2)\mathrm{trace}(\tilde{\mathbf{J}}_1^{-1}\tilde{\mathbf{J}}_0)] - \gamma_n^2(p-1)^2/2}{2n^2\eta_n^2} \times (1+o(1)), \qquad \text{(B9)}$$

as $n \to \infty$.

We prove (B9) first. Let $\boldsymbol{b}' = (b_1, \ldots, b_{p-1})^T$, $\boldsymbol{b}_n'^* = (b_{n,1}'^*, \ldots, b_{n,p-1}'^*)^T$, $\boldsymbol{u}_0' = (u_{0,1} - u_{0,p}, \ldots, u_{0,p-1} - u_{0,p})^T$, $\boldsymbol{x}_i' = (x_{i,1}, \ldots, x_{i,p-1})^T$, $\tilde{\boldsymbol{b}} = \sqrt{n}(\boldsymbol{b}' - \boldsymbol{b}_n'^*)$, and $\tilde{u}_0 = (\frac{\tilde{\mathbf{J}}_0}{n} + \eta_n\tilde{\mathbf{J}}_1 + \frac{\tilde{\mathbf{J}}_2}{n})^{-1}\boldsymbol{u}_0'$, where $(u_{0,1}, \ldots, u_{0,p})^T = \boldsymbol{u}_0$. Observe that

$$\begin{aligned}
&\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\exp\Big[-\frac{n}{2}(\boldsymbol{b}-\boldsymbol{b}_n^*)^T\left(\frac{\mathbf{J}_0}{n}+\eta_n\mathbf{J}_1+\frac{\mathbf{J}_2}{n}\right)(\boldsymbol{b}-\boldsymbol{b}_n^*) \\
&\quad + \boldsymbol{u}_0^T(\boldsymbol{b}-\boldsymbol{b}_n^*)\Big]\mathrm{d}\boldsymbol{b} \\
&= \int_{\|\boldsymbol{b}-\boldsymbol{b}_n^*\|\le\zeta_n,\boldsymbol{b}\in B}\exp\Big[-\frac{n}{2}(\boldsymbol{b}'-\boldsymbol{b}_n'^*)^T\left(\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1+\frac{\tilde{\mathbf{J}}_2}{n}\right) \\
&\quad \times(\boldsymbol{b}'-\boldsymbol{b}_n'^*) + \boldsymbol{u}_0'^T(\boldsymbol{b}'-\boldsymbol{b}_n'^*)\Big]\mathrm{d}\boldsymbol{b}' \\
&= \frac{1}{\sqrt{n}}\int_{\mathbb{R}^{p-1}}\exp\left[-\frac{1}{2}\tilde{\boldsymbol{b}}^T\left(\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1+\frac{\tilde{\mathbf{J}}_2}{n}\right)\tilde{\boldsymbol{b}} + \frac{1}{\sqrt{n}}\boldsymbol{u}_0'^T\tilde{\boldsymbol{b}}\right]\mathrm{d}\tilde{\boldsymbol{b}} \\
&= \frac{1}{\sqrt{n}}e^{-\frac{\boldsymbol{u}_0'^T\left(\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1+\frac{\tilde{\mathbf{J}}_2}{n}\right)^{-1}\boldsymbol{u}_0'}{2n}}\int_{\mathbb{R}^{p-1}}\exp\Big[-\frac{1}{2}\left(\tilde{\boldsymbol{b}}-\frac{\tilde{\boldsymbol{u}}_0}{\sqrt{n}}\right)^T \\
&\quad \times\left(\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1+\frac{\tilde{\mathbf{J}}_2}{n}\right)\left(\tilde{\boldsymbol{b}}-\frac{\tilde{\boldsymbol{u}}_0}{\sqrt{n}}\right)\Big] \\
&= \frac{1}{\sqrt{n}}e^{-\frac{\boldsymbol{u}_0'^T\left(\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1+\frac{\tilde{\mathbf{J}}_2}{n}\right)^{-1}\boldsymbol{u}_0'}{2n}}(2\pi)^{-\frac{p-1}{2}}\left|\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1+\frac{\tilde{J}_2}{n}\right|^{-\frac{1}{2}},
\end{aligned}$$

where

$$\tilde{\mathbf{J}}_0 = \frac{(\boldsymbol{x}_{n+1}'-x_{n+1,p}\mathbf{1}_p)(\boldsymbol{x}_{n+1}'-x_{n+1,p}\mathbf{1}_p)^T}{(\boldsymbol{x}_{n+1}^T\boldsymbol{b}_n^*)^2};$$

$$\tilde{\mathbf{J}}_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{(\boldsymbol{x}_i'-x_{i,p}\mathbf{1}_p)(\boldsymbol{x}_i'-x_{i,p}\mathbf{1}_p)^T}{(\boldsymbol{x}_i^T\boldsymbol{b}_n^*)^2};\ \tilde{\mathbf{J}}_2 = \gamma_n\tilde{\mathbf{J}}_1.$$

It follows that

$$\begin{aligned}
U_n &= \frac{\exp\left(-\frac{\boldsymbol{u}_0'^T\left(\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1+\frac{\tilde{\mathbf{J}}_2}{n}\right)^{-1}\boldsymbol{u}_0'}{2n}\right)}{\exp\left(-\frac{\boldsymbol{u}_0'^T(\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1)^{-1}\boldsymbol{u}_0'}{2n}\right)} \\
&\quad\times\frac{\left|\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1\right|^{0.5}\left|\eta_n\tilde{\mathbf{J}}_1+\frac{\tilde{\mathbf{J}}_2}{n}\right|^{0.5}}{\left|\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1+\frac{\tilde{\mathbf{J}}_2}{n}\right|^{0.5}|\eta_n\tilde{\mathbf{J}}_1|^{0.5}} - 1 \\
&= (1+A)(1+B)(1+C) - 1 = ABC + AB + BC \\
&\quad + CA + A + B + C, \qquad \text{(B10)}
\end{aligned}$$

where

$$A = \frac{\exp(-\frac{\boldsymbol{u}_0'^T(\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1+\frac{\tilde{\mathbf{J}}_2}{n})^{-1}\boldsymbol{u}_0'}{2n})}{\exp(-\frac{\boldsymbol{u}_0'^T(\frac{\tilde{\mathbf{J}}_0}{n}+\eta_n\tilde{\mathbf{J}}_1)^{-1}\boldsymbol{u}_0'}{2n})} - 1,$$

$$B = \frac{|\frac{\tilde{\mathbf{J}}_0}{n} + \eta_n \tilde{\mathbf{J}}_1|^{0.5}}{|\frac{\tilde{\mathbf{J}}_0}{n} + \eta_n \tilde{\mathbf{J}}_1 + \frac{\tilde{\mathbf{J}}_2}{n}|^{0.5}} - 1, \text{ and } C = \frac{|\eta_n \tilde{\mathbf{J}}_1 + \frac{\tilde{\mathbf{J}}_2}{n}|^{0.5}}{|\eta_n \tilde{\mathbf{J}}_1|^{0.5}} - 1.$$

Note that

$$\tilde{\mathbf{J}}_1 = \frac{-\partial^2 \log W_n(\boldsymbol{b})}{\partial \boldsymbol{b}' \partial \boldsymbol{b}'^T} = (\mathbf{I}_{p-1}, -\mathbf{1}_{p-1}) \times \frac{-\partial^2 \log W_n(\boldsymbol{b})}{\partial \boldsymbol{b} \partial \boldsymbol{b}^T} (\mathbf{I}_{p-1}, -\mathbf{1}_{p-1})^T,$$

and from Theorem 2 of Gaivoronski and Stella (2000), $\tilde{\mathbf{J}}_1$ is positive definite and

$$\lambda_{\min}(\tilde{\mathbf{J}}_1) \geq \delta > 0.$$

From Taylor expansion, we have

$$\left(1 + \frac{ar + br^2 + O(r^3)}{1 + a'r + b'r^2 + O(r^3)}\right)^{0.5} = 1 + \frac{ar}{2} + \left(\frac{b}{2} - \frac{a^2}{8}\right) r^2 + o(r^2)$$

for constants a, b, a', b', and

$$|\mathbf{I} + \mathbf{X}r| = 1 + \text{trace}(\mathbf{X})r + \left[\text{trace}(\mathbf{X})^2 - \text{trace}(\mathbf{X}^2)\right] \times \frac{r^2}{2} + o(r^2), \ (\mathbf{I} + r\mathbf{X})^{-1} = \mathbf{I} - r\mathbf{X} + o(r),$$

for an invertible matrix X, and identity matrix $\mathbf{I}$, as scalar $r \to 0$. As a result, we have the following expansions.

$$C = \left(1 + \frac{\gamma_n}{n\eta_n}\right)^{\frac{p-1}{2}} - 1 \sim \frac{p-1}{2}\frac{\gamma_n}{n\eta_n} + \frac{(p-1)(p-3)}{8}\frac{\gamma_n^2}{(n\eta_n)^2} + o\left(\frac{1}{(n\eta_n)^2}\right)$$

$$B \sim \left(1 - \frac{\frac{(p-1)\gamma_n}{n\eta_n} + \frac{1}{2n^2\eta_n^2}M}{1 + \frac{\text{trace}(\mathbf{J}_3)+(p-1)\gamma_n}{n\eta_n} + \frac{1}{2n^2\eta_n^2}\left[\text{trace}^2(\mathbf{J}_3) - \text{trace}(\mathbf{J}_3^2) + M\right]}\right)^{0.5} - 1$$
$$\sim -\frac{(p-1)\gamma_n}{2n\eta_n} - \left(\frac{M}{4} + \frac{(p-1)^2\gamma_n^2}{8}\right)\frac{1}{n^2\eta_n^2} + o\left(\frac{1}{(n\eta_n)^2}\right),$$

$$A \sim \boldsymbol{u}_0'^T\left[\left(\frac{\tilde{\mathbf{J}}_0}{n} + \eta_n\tilde{\mathbf{J}}_1\right)^{-1} - \left(\frac{\tilde{\mathbf{J}}_0}{n} + \eta_n\tilde{\mathbf{J}}_1 + \frac{\tilde{\mathbf{J}}_2}{n}\right)^{-1}\right]\boldsymbol{u}_0'/2n$$
$$= \boldsymbol{u}_0'^T \frac{\left[\mathbf{I} - \left(\mathbf{I} + \left(\frac{\tilde{\mathbf{J}}_0}{n\eta_n} + \tilde{\mathbf{J}}_1\right)^{-1}\tilde{\mathbf{J}}_1\frac{\gamma_n}{n\eta_n}\right)^{-1}\right]\left[\frac{\tilde{\mathbf{J}}_0}{n\eta_n} + \tilde{\mathbf{J}}_1\right]^{-1}}{2n\eta_n}\boldsymbol{u}_0'$$
$$= \boldsymbol{u}_0'^T \frac{\left[\frac{\tilde{\mathbf{J}}_0}{n\eta_n} + \tilde{\mathbf{J}}_1\right]^{-1}\tilde{\mathbf{J}}_1\frac{\gamma_n}{n\eta_n}\left[\frac{\tilde{\mathbf{J}}_0}{n\eta_n} + \tilde{\mathbf{J}}_1\right]^{-1}}{2n\eta_n}\boldsymbol{u}_0$$
$$\sim \frac{\gamma_n\boldsymbol{u}_0'^T\tilde{\mathbf{J}}_1^{-1}\boldsymbol{u}_0'}{2n^2\eta_n^2} + o\left(\frac{1}{(n\eta_n)^2}\right),$$

where $M = \gamma_n^2(p-1)(p-2) + 2\gamma_n\text{trace}(\mathbf{J}_3)(p-2)$ and $\mathbf{J}_3 = \tilde{\mathbf{J}}_0\tilde{\mathbf{J}}_1^{-1}$. Continue with (B10), from Condition (A.6),

$$U_n \sim ABC + AB + BC + CA + A + B + C \sim \frac{\gamma_n[\boldsymbol{u}_0'^T\tilde{\mathbf{J}}_1^{-1}\boldsymbol{u}_0' - (p-2)\text{trace}(\tilde{\mathbf{J}}_1^{-1}\tilde{\mathbf{J}}_0)] - \gamma_n^2(p-1)^2/2}{2n^2\eta_n^2}.$$

Thus (B9) holds.

Next, we prove (B8). It follows from the integral mean value theorem that

$$\frac{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_{n+1})W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b}}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_{n+1})W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b} \times \int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}} - 1$$
$$= \frac{\exp(O(n\eta_n\zeta_n^7))\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\exp(V_1 + \eta_n V_2 + V_3)\mathrm{d}\boldsymbol{b}\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\exp(\eta_n V_2)\mathrm{d}\boldsymbol{b}}{\exp(O(n\eta_n\zeta_n^7))\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\exp(V_1 + \eta_n V_2)\mathrm{d}\boldsymbol{b} \times \int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\exp(V_3 + \eta_n V_2)\mathrm{d}\boldsymbol{b}} - 1$$
$$\sim \frac{\exp(O(n\eta_n\zeta_n^7))}{\exp(O(n\eta_n\zeta_n^7))} - 1 + D - 1,$$

where

$$D = \frac{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\exp(V_1 + \eta_n V_2 + V_3)\mathrm{d}\boldsymbol{b}\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\exp(V_2)\mathrm{d}\boldsymbol{b}}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\exp(V_1 + \eta_n V_2)\mathrm{d}\boldsymbol{b} \times \int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\exp(V_3 + \eta_n V_2)\mathrm{d}\boldsymbol{b}}.$$

From Taylor expansion, if $\alpha = 7$, or $(\log n)^7/n = o(\eta_n)$, we have

$$\frac{\exp(O(n\eta_n\zeta_n^7))}{\exp(O(n\eta_n\zeta_n^7))} - 1 = O(n\eta_n\zeta_n^7) = O\left(\frac{(\log n)^{3.5}}{(n\eta_n)^{2.5}}\right) = o\left(\frac{1}{(n\eta_n)^2}\right).$$

In addition,

$$D - 1 \sim \frac{\gamma_n[\boldsymbol{u}_0'^T\tilde{\mathbf{J}}_1^{-1}\boldsymbol{u}_0' - (p-2)\text{trace}(\tilde{\mathbf{J}}_1^{-1}\tilde{\mathbf{J}}_0)] - \gamma_n^2(p-1)^2/2}{2n^2\eta_n^2},$$

based on the fact that, for any kind of scalar sequences (a_n), (b_n), (c_n), and (d_n),

$$\frac{a_n + c_n}{b_n + d_n} - 1 \sim \frac{a_n}{b_n} - 1$$

holds, as long as $c_n - d_n = o\big(b_n(\frac{a_n}{b_n} - 1)\big)$, where $c_n = o(a_n)$, $d_n = o(b_n)$. Then (B8) holds.

If $\alpha \in (1, 7)$, take the integer

$$K = \min\left\{k \in \mathbb{Z} \mid k \geq \frac{6\alpha}{\alpha - 1}\right\}.$$

Using Taylor expansion of each term in (B8) up to the $(K-1)$th order, we have

$$\frac{\exp(O(n\zeta_n^K))}{\exp(O(n\zeta_n^K))} - 1 = O(n\zeta_n^K) = o\left(\frac{1}{n^2\eta_n^2}\right),$$

and (B8) can be proved along similar lines. In conclusion both (B8) and (B9) hold.

(II) In this part, we prove the first statement of part (b). Choose any

$$\epsilon \in \left(\frac{3}{2p+6}, 0.5\right),$$

and η_n and ζ_n the same as in Part (I).

Next similar to (B3) of lemma 2.1, using Condition (A.5)

$$\frac{1}{n}\log W_n(\boldsymbol{b}_n^*) \in [c_4, c_5],$$

we have

$$\frac{\int_{B\backslash B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_{n+1})W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_{n+1})W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b}} = nO\left(n^{-\frac{p\epsilon}{1-2\epsilon}}\right)$$

and

$$\frac{\int_{B\backslash B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b}}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}} = \frac{1}{n}O(n^{-\frac{p\epsilon}{1-2\epsilon}}).$$

Then we have,

$$\frac{\int_{B\backslash B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_{n+1})W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}\int_{\mathrm{B}\backslash\mathrm{B}_{\zeta_n}(\boldsymbol{b}_n^*)}\mathrm{W}_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b}}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_{n+1})W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b} \times \int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}} = O(n^{-\frac{2p\epsilon}{1-2\epsilon}}) = o\left(\frac{1}{n^2\eta_n^2}\right),$$

and

$$\frac{\int_{B\backslash B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_{n+1})W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b}\int_{B\backslash B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}}{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_{n+1})W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b} \times \int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}} = O(n^{-\frac{2p\epsilon}{1-2\epsilon}}) = o\left(\frac{1}{n^2\eta_n^2}\right).$$

Hence,

$$
\begin{aligned}
&\frac{\mathrm{Cov}_{B\setminus B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta},\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)}{\mathrm{Cov}_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta},\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)}\\
&=\frac{\mathrm{E}_{B\setminus B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta}\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)-\mathrm{E}_{B\setminus B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\boldsymbol{b}_{\eta}\big)\mathrm{E}_{B\setminus B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)}{\mathrm{E}_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta}\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)-\mathrm{E}_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta}\big)\mathrm{E}_{B\setminus B_{n-1,\zeta_n}(\boldsymbol{b}_n^*)}\big(\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)}\\
&=\frac{\dfrac{\mathrm{E}_{B\setminus B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta}\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)}{\mathrm{E}_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta}\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)}-\dfrac{\mathrm{E}_{B\setminus B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta}\big)\mathrm{E}_{B\setminus B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)}{\mathrm{E}_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta}\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)}}{1-\dfrac{\mathrm{E}_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta}\big)\mathrm{E}_{B\setminus B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)}{\mathrm{E}_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta}\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)}}=0.
\end{aligned}
$$

In the end, there exists $N>0$, for any $n>N$,

$$
\begin{aligned}
&\mathrm{Cov}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta},\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)\\
&=\mathrm{Cov}_{B_{\zeta_n}(\boldsymbol{b}_n^*)}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta},\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)\\
&=\frac{\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}(\boldsymbol{b}^T\boldsymbol{x}_n)W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b}\times\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\log W_n(\boldsymbol{b})\mathrm{d}\boldsymbol{b}}{[\int_{B_{\zeta_n}(\boldsymbol{b}_n^*)}W_n^{\eta_n}(\boldsymbol{b})\mathrm{d}\boldsymbol{b}]^2}\\
&\times\frac{\gamma_n[\boldsymbol{u}_0'^T\tilde{\mathbf{J}}_1^{-1}\boldsymbol{u}_0'-(p-2)\mathrm{trace}(\tilde{\mathbf{J}}_1^{-1}\tilde{\mathbf{J}}_0)]-\gamma_n^2(p-1)^2/2}{2n^2\eta_n^2}\\
&\times(1+o(1)),
\end{aligned}
$$

which keeps its sign uniformly for $\eta_n\in[\eta_{1,n},+\infty)$. Part (a) of the theorem is proved.

(III) We now prove the second statement of Part (b) when Condition (A.7) holds. Using Taylor expansion, as $n\eta\to 0$,

$$W_n^{\eta}(\boldsymbol{b})=\exp\left[\frac{1}{n}\log W_n(\boldsymbol{b})(n\eta)\right]\sim 1+\eta\log W_n(\boldsymbol{b}).$$

Asymptotically,

$$
\begin{aligned}
&\int W_n^{\eta}(\boldsymbol{b})\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b})\\
&\sim\int\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b})+(n\eta)\int\frac{1}{n}\log W_n(\boldsymbol{b})\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b})\\
&\le\int\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b})+(n\eta)\int\frac{1}{n}\log W_n(\boldsymbol{b}_n^*)\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b})\\
&\le\int\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b})+(n\eta)c_5\int\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b}).
\end{aligned}
$$

Similarly, for $k=0,1$,

$$
\begin{aligned}
&\int\log W_n(\boldsymbol{b})W_n^{\eta}(\boldsymbol{b})(\boldsymbol{x}_{n+1}^T\boldsymbol{b})^k\pi(\mathrm{d}\boldsymbol{b})\\
&\quad\sim\int\log W_n(\boldsymbol{b})(\boldsymbol{x}_{n+1}^T\boldsymbol{b})^k\pi(\mathrm{d}\boldsymbol{b})+O(n\eta).
\end{aligned}
$$

Then

$$
\begin{aligned}
&\int\log W_n(\boldsymbol{b})W_n^{\eta}(\boldsymbol{b})\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b})-\int W_n^{\eta}(\boldsymbol{b})\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b})\\
&\quad\times\int\log W_n(\boldsymbol{b})W_n^{\eta}(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})\\
&=\int\log W_n(\boldsymbol{b})\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b})-\int\boldsymbol{x}_{n+1}^T\boldsymbol{b}\pi(\mathrm{d}\boldsymbol{b})\\
&\quad\times\int\log W_n(\boldsymbol{b})\pi(\mathrm{d}\boldsymbol{b})+o(1).
\end{aligned}
$$

With Condition (A.5), it follows that,

$$\mathrm{Cov}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{\eta},\log W_{n-1}(\mathfrak{b}_{n-1}^{\eta})\big)=\mathrm{Cov}\big(\boldsymbol{x}_n^T\mathfrak{b}_{n-1}^{0},\log W_{n-1}(\mathfrak{b}_{n-1}^{0})\big)\big(1+o(1)\big).$$

We get the conclusion from $\mathfrak{b}_{n-1}^0=\boldsymbol{b}^{(0)}$ and Condition (A.7). □

Proof of Theorem 3.5 Write

$$
\begin{aligned}
&\mathrm{E}\left[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\tilde{\eta}_i)}|\mathcal{F}_{i-1},i\in\mathcal{C}_{0,0}\right]\\
&\quad=\mathrm{E}\left[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)};\kappa(i)=1|\mathcal{F}_{i-1},i\in\mathcal{C}_{0,0}\right]\\
&\qquad\times\mathrm{P}\big(\kappa(i)=1|\mathcal{F}_{i-1},i\in\mathcal{C}_{0,0}\big)\\
&\qquad+\mathrm{E}[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(0)};\kappa(i)=2|\mathcal{F}_{i-1},i\in\mathcal{C}_{0,0}]\\
&\qquad\times\mathrm{P}\big(\kappa(i)=2|\mathcal{F}_{i-1},i\in\mathcal{C}_{0,0}\big)\\
&\quad=0,
\end{aligned}
\tag{B11}
$$

and

$$
\begin{aligned}
&\mathrm{E}[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\tilde{\eta}_i)}|\mathcal{F}_{i-1},i\in\mathcal{C}_{\infty,\infty}]\\
&\quad=\mathrm{E}[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)};\kappa(i)=1|\mathcal{F}_{i-1},i\in\mathcal{C}_{\infty,\infty}]\\
&\qquad\times\mathrm{P}\big(\kappa(i)=1|\mathcal{F}_{i-1},i\in\mathcal{C}_{\infty,\infty}\big)\\
&\qquad+\mathrm{E}[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(0)};\kappa(i)=2|\mathcal{F}_{i-1},i\in\mathcal{C}_{\infty,\infty}]\\
&\qquad\times\mathrm{P}\big(\kappa(i)=2|\mathcal{F}_{i-1},i\in\mathcal{C}_{\infty,\infty}\big)\\
&\quad=\mathrm{E}[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\infty)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(0)};\kappa(i)=2|\mathcal{F}_{i-1},i\in\mathcal{C}_{\infty,\infty}]\\
&\qquad\times\mathrm{P}\big(\kappa(i)=2|\mathcal{F}_{i-1},i\in\mathcal{C}_{\infty,\infty}\big)\\
&\quad\le u\mathrm{E}[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\infty)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(0)}|\mathcal{F}_{i-1},i\in\mathcal{C}_{\infty,\infty}].
\end{aligned}
\tag{B12}
$$

Similar to (B11) and (B12),

$$
\begin{aligned}
&\mathrm{E}\left[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\tilde{\eta}_i)}|\mathcal{F}_{i-1},i\in\mathcal{C}_{0,\infty}\right]\\
&\quad=\mathrm{E}\left[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\infty)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(0)}|\mathcal{F}_{i-1},i\in\mathcal{C}_{0,\infty}\right],
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{E}\left[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\tilde{\eta}_i)}|\mathcal{F}_{i-1},i\in\mathcal{C}_{\infty,0}\right]\\
&\quad\le u\mathrm{E}\left[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(0)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\infty)}|\mathcal{F}_{i-1},i\in\mathcal{C}_{\infty,0}\right].
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\log W^*_{n|(\eta^*)_n}-\log\tilde{W}_{n|(\eta^*)_n}\\
&=\sum_{i=1}^{n}\mathrm{E}\left[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\eta_i^*)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\tilde{\eta}_i)}|\mathcal{F}_{i-1},\eta_i^*\right]\\
&\le u\left\{\sum_{i\in\mathcal{C}_{\infty,\infty}}\mathrm{E}\left[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\infty)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(0)}|\mathcal{F}_{i-1},\eta_i^*=\infty\right]\right.\\
&\left.+\sum_{i\in\mathcal{C}_{\infty,0}}\mathrm{E}\left[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(0)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\infty)}|\mathcal{F}_{i-1},\eta_i^*=0\right]\right\}\\
&+\sum_{i\in\mathcal{C}_{0,\infty}}\mathrm{E}\left[\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(\infty)}-\log\boldsymbol{x}_i^T\boldsymbol{b}_i^{(0)}|\mathcal{F}_{i-1},\eta_i^*=\infty\right]\\
&=u(\Upsilon_{\infty,\infty}+\Upsilon_{0,\infty})+\Upsilon_{\infty,0}.
\end{aligned}
\tag{B13}
$$

Again, analogous to the derivation of (B11) and (B12), we have

$$\mathrm{E}\left[\log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(\tilde{\eta}_i)} - \log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(0)} | \mathcal{F}_{i-1}, i \in \mathcal{C}_{j,1}\right] = 0,\ j = 1, 2,$$
$$\mathrm{E}\left[\log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(\tilde{\eta}_i)} - \log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(0)} | \mathcal{F}_{i-1}, i \in \mathcal{C}_{\infty,\infty}\right] \geq (1-u)$$
$$\times\, \mathrm{E}\left[\log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(\infty)} - \log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(0)} | \mathcal{F}_{i-1}, i \in \mathcal{C}_{\infty,\infty}\right],$$
$$\mathrm{E}\left[\log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(\tilde{\eta}_i)} - \log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(0)} | \mathcal{F}_{i-1}, i \in \mathcal{C}_{\infty,0}\right] \geq -u$$
$$\mathrm{E}\left[\log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(0)} - \log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(\infty)} | \mathcal{F}_{i-1}, i \in \mathcal{C}_{\infty,0}\right].$$

Hence,

$$\begin{aligned}
&\log \tilde{W}_{n|(\eta^*)_n} - \log W^{(0)}_{n|(\eta^*)_n} \\
&= \sum_{i=1}^{n} \mathrm{E}\left[\log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(\tilde{\eta}_i)} - \log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(0)} | \mathcal{F}_{i-1}, \eta_i^*\right] \\
&\geq (1-u) \sum_{i \in \mathcal{C}_{\infty,\infty}} \mathrm{E}\left[\log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(\infty)} - \log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(0)} | \mathcal{F}_{i-1}, \eta_i^* = \infty\right] \\
&\quad - u \sum_{i \in \mathcal{C}_{\infty,0}} \mathrm{E}\left[\log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(0)} - \log \boldsymbol{x}_i^T \boldsymbol{b}_i^{(\infty)} | \mathcal{F}_{i-1}, \eta_i^* = 0\right] \\
&= \Upsilon_{\infty,\infty} - u(\Upsilon_{\infty,0} + \Upsilon_{\infty,\infty}). \qquad \text{(B14)}
\end{aligned}$$

Combining (B13) and (B14) leads to the desired conclusion. □