Universal Portfolios For Target Classes Having Continuous Form Dependence On Side Information

A Dissertation Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Jason Earle Cross

Dissertation Director: Andrew R. Barron

May 1999

UMI Number: 9931012

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Abstract

Universal Portfolios For Target Classes Having Continuous Form Dependence On Side Information

Jason Earle Cross

1999

This thesis generalizes and extends some of the concepts of universal portfolios as introduced by Cover (1991). We begin with an abstract framework, generalizing the concept of target class beyond constant rebalanced portfolios to include potential classes having continuous form dependence on side information. An analogy of Cover's universal portfolio in discrete time is addressed by extending results to continuous time where we show the existence of an easily computable, continuously updated, universal procedure for linearly parameterized classes of portfolios. Finally, to reconcile ease of computation with applicability we discretize the above procedure and analyze it in near continuous time. Given an appropriate schedule of increasingly frequent rebalances, we propose that this discretized portfolio remains universal with respect to its continuously traded target class.

I dedicate this dissertation to my family, both old and new...

Contents

List of Figures

List of Tables

Acknowledgements

I am deeply indebted to many people for their part in seeing this work to fruition.

First and foremost I must acknowledge my advisor, Dr. Andrew Barron, who has been an unparalleled source of inspiration, encouragement, and leadership throughout this process. His frequent insights and guidance into my research have been invaluable.

I would also like to thank Dr. Nicolas Hengartner for his generous support and guidance throughout my years at Yale. I am especially grateful for his help during the prospectus phase of my research and for the many helpful suggestions he has contributed throughout this process.

I also thank Dr. Marten Wegkamp for reading this thesis and for sharing helpful insights that greatly improved early versions of this dissertation.

To Jay Emerson, John Li, Brendon Murphy, and Drew Carter, I thank you for your many years of fellowship and friendship and for making Yale an enjoyable place to study.

I also thank my parents for their belief and support for which I am truly indebted.

Finally, I thank Anne Boyle, my bride to be, for her patience, caring, and support during the writing of this thesis. To me, you continue to be a constant reminder of what is important in life.

Chapter 1

Introduction and Background

1.1 Preliminaries

Suppose an investor is faced with the challenge of reallocating wealth among m stocks and cash at the start of n consecutive trading periods indexed by $i \in \{1, \ldots, n\}$. The allocation at the beginning of period i is represented by a portfolio vector $b_i = (b_{i,0}, b_{i,1}, \ldots, b_{i,m})$. Here $b_{i,0}$ represents the proportion of wealth in cash and $b_{i,j}$ represents the proportion of wealth in stock j at the start of period i. We assume that the portfolio vector satisfies two constraints, namely $\sum_{j=0}^{m} b_{i,j} = 1$ and $b_{i,j} \geq 0$ for $0 \leq j \leq m$. The first constraint insures that the portfolio is *self-financing*, i.e. that there is no inflow or outflow of capital required to invest in the portfolio. The second constraint insures that short selling stock ($b_{i,j} < 0$ for $1 \leq j \leq m$) and/or purchasing stock on margin ($b_{0,j} < 0$) is prohibited.

For each investment period we denote the price of stock j at the end of period i by $P_{i,j}$. Given these prices we define the *wealth relative*, $X_{i,j} = P_{i,j}/P_{i-1,j}$, to be the ratio of the price of stock j at the end of period i to that at the beginning of period i. Collectively we write the vector of wealth relatives as,

$$X_i = (1, X_{i1}, \ldots, X_{im}),$$

with the understanding that the first component, the wealth relative of cash, is always 1. We can think of wealth relatives as the factor by which the value of a stock increases over one period. For example, a wealth relative of 1.1 corresponds to a 10% increase in value. We can also talk of the wealth relative generated by a portfolio. The quantity $b_i'X_i$ represents the factor by which b_i increases wealth over period i. Using this interpretation and the definitions above we see that if we start with an initial wealth of W_0 and use a sequence of portfolios $\{b_i\}_{i=1}^n$, by the end of n time periods our wealth is,

$$W_n = W_0 \prod_{i=1}^{n} b_i' X_i.$$

Thus the wealth is simply a product of the individual portfolio relatives $b_i' X_i$ achieved each period.

1.2 Log-Optimal Investment

Suppose that the sequence of wealth relatives $X_1, X_2, \ldots, X_n$ is generated according to some known probability distribution P. Given P, we want to use a portfolio sequence $\{b_i\}_{i=1}^n$ that maximizes wealth, W_n, in some sense. But since W_n is a random variable representing a terminal *distribution* of wealth, it is not immediately clear what we mean by this maximization. What does it mean to maximize a random variable and its distribution? Obviously we need a way of ranking distributions according to their desirability. This is done by using a functional of W_n that collapses the relative merits of each distribution to a single number. Then the problem of "maximizing" W_n is reduced to choosing the portfolio sequence $\{b_i\}_{i=1}^n$ that maximizes our chosen functional.

But which functional should we maximize? What criterion should we use? There are many options. However, we feel the best argument can be made for choosing sequences $\{b_i\}_{i=1}^n$ that maximize expected log wealth, $E \log W_n$. Arguments favoring this criterion are strong. We are particularly motivated by the works of Kelly [9], Latane [12], Breiman [4], Bell and Cover [3], and Algoet and Cover [2] who show the growth optimality of this criterion under many circumstances.

To understand the optimality of this criterion we first consider the behavior of W_n when using a *constant rebalanced portfolio* b. A constant rebalanced portfolio (CPR) is simply a portfolio that resets wealth allocation at the end of each period to that used at the start of the period. That is, a CRP uses the same wealth proportions each period. This is not to be confused with a buy and hold strategy for which no trading occurs. Contrary to their name, CRP's require a lot of trading. It is usually necessary to buy and sell appropriate amounts of each stock at the end of each period to insure that the proportions of wealth return to their initial values.

The wealth achieved by CRP b has the simple expression,

$$W_n = W_0 \prod_{i=1}^{n} b' X_i.$$

Suppose that an investor uses a CRP in a market of i.i.d. wealth relatives X. In this case, we see from the form of W_n that the wealth becomes a product of i.i.d. factors. In turn $\log W_n$ becomes a sum of i.i.d. terms. Hence the behavior of $\log W_n$ is governed by the law of large numbers and we can infer that,

$$\frac{1}{n}\log W_n = E\log bX + Y_n, \quad \text{where } Y_n \to 0 \text{ w. p. 1.}$$

We now define the *empirical growth rate* of wealth as, $R_n \equiv \frac{1}{n}\log W_n/W_0$. The growth rate gives a measure of how quickly the wealth grows exponentially. In terms of its empirical growth rate we can write W_n as,

$$W_n = W_0 \exp\{n(R_n)\}.$$

By definition, W_n grows exponentially according to its growth rate R_n regardless of stochastic assumptions. However in our i.i.d. market we see that, asymptotically, the growth rate converges to $E\log bX$.

$$\lim_{n\to\infty} R_n = \lim_{n\to\infty}\frac{1}{n}\log W_n - \lim_{n\to\infty}\frac{1}{n}\log W_0 = E\log bX$$

Thus the growth rate is asymptotically maximized when using portfolio b such that $E\log bX$ is maximized. But in the i.i.d. case this is equivalent to maximizing $E\log W_n$ since,

$$E\log W_n = \log W_0 + nE\log bX.$$

Therefore by choosing a CRP b to maximize $E\log W_n$ we effectively maximize the asymptotic growth rate and hence wealth. More importantly, by using such a portfolio we are guaranteed with almost sure probability to beat the asymptotic growth rate (and wealth) of any other constant rebalanced portfolio.

Now suppose that we are allowed to invest in any type of causal portfolio sequence. What is the optimal strategy now? Breiman [4] shows that it is still optimal to select the CRP maximizing log wealth. To be precise, if W_n^* is the wealth achieved by maximizing $E\log W_n$ and if W_n is the wealth achieved by any other portfolio sequence then,

$$\lim_{n\to\infty}\sup_{\{X_i\}_{i=1}^n}\frac{1}{n}\log\frac{W_n}{W_n^*} \le 0 \tag{1.1}$$

with probability 1. Thus asymptotically the log optimal strategy does at least as well as any other strategy up to first order in the exponent.

Even more strikingly, similar results can been shown for the log optimal strategy when wealth relatives are assumed to come from some known but not necessarily i.i.d. process. Algoet and Cover [2] show that the wealth W_n^* achieved by the log optimal sequence,

$$b_i(X_{i-1},\dots,X_1) = \arg\max_b E\left[\log\left(b'X_i\right)|X_{i-1},\dots,X_1\right]$$

also satisfies (1.1) when compared to the wealth of any other sequence. If we further assume that the process generating the wealth relatives is stationary ergodic, the asymptotic growth rate is equal to,

$$\lim_{n\to\infty} E\log\left(X_1' b(X_0, X_{-1}, \dots, X_{-n})\right).$$

Given its optimal properties, we feel the case for maximizing expected log wealth is strong. However we must point out that the acceptance of the log wealth criterion is not universal. Other paradigms of portfolio selection exist. For instance, much of the previous work in portfolio theory has centered on the mean-variance approach pioneered by Markowitz [15]. Here it is argued that an investor should choose a sequence of portfolios that maximizes the first moment of W_n subject to a constraint on variance. In this setting, variance becomes a proxy for risk and the investor tries to maximize expected return for a given level of risk. This is the basis for the Sharpe-Markowitz theory of investment. The book by Sharpe [21] provides an excellent introduction to this topic.

More generally, the traditional view of finance has been that an investor should choose a portfolio according to a subjective utility function that quantifies the investor's preference for money and risk. A reasonable utility function is both increasing and concave. Increasing, because more money is better, and concave because investors are risk adverse. In this context, the log criterion becomes one of many possibilities. The choice of utility function depends only on the personal preferences of the investor. It is completely subjective. However, we believe adherence to a subjective utility function is questionable if the investment horizon is long. For very large n we are almost surely guaranteed to make more money by maximizing log wealth than by maximizing any other utility function. Thus if more money is truly better, it would seem unreasonable to use anything other than the log criterion in the long run. A vigorous debate over the merits of this viewpoint has been waged for years. For criticisms of the log-optimal approach, the reader is directed to the papers of Samuelson [20] and Ophir [18], [19]. Counterpoints to these criticisms are given by Latane [13], [14]. Excellent discussions of the merits of log-optimal investment and a summary of its important properties are given by both Cover and Thomas [7] and Larson [11].

1.3 Universal Portfolios

1.3.1 Definition

We have argued that if the asymptotic growth rate of wealth is to be maximized, we should select the portfolio sequence that maximizes expected log wealth. Of course this can only be done if the process governing wealth relatives is known. In actuality this is rarely the case. Usually we can assume very little about the behavior of wealth relatives, especially in the long run.

What we now seek is a way of achieving growth optimal wealth without using any stochastic assumptions whatsoever. We henceforth interpret $X_1, X_2, \ldots, X_n$ as an arbitrary sequence of vectors in $\mathbf{R}_+^{m+1}$ and instead of optimizing the asymptotic growth rate (and hence wealth) in an almost sure sense we now aim to maximize these quantities uniformly over all sequences.

To see how this might be done, we consider a scenario where we are allowed to choose any portfolio sequence from a set of sequences B. If we were given foresight of $X_1, X_2, \ldots, X_n$ we could choose an optimal portfolio sequence, $\{b_i^*\}_{i=1}^n$, from the set B that maximizes the empirical growth rate $R_n \equiv \frac{1}{n} \log W_n/W_0$. Define R_n^* to be the maximal rate achieved by this optimal sequence $\{b_i^*\}_{i=1}^n$. What we want is to find another portfolio sequence $\left\{\widehat{b}_i\right\}_{i=1}^n$ not necessarily in B and *not* dependent on future knowledge of $X_1, X_2, \ldots, X_n$ that has a growth rate $\widehat{R}_n$ asymptotically equal to R_n^*. We cannot expect that such a portfolio always exists but if one does exist it means that we can achieve some fairly impressive results. Essentially, such a portfolio will achieve, without future knowledge of prices, almost the same wealth as if you were given future knowledge of prices and then allowed to act according to any portfolio sequence in B. We call $\left\{\widehat{b}_i\right\}_{i=1}^n$ a *universal portfolio* because it can achieve this feat *universally* over all sequences of wealth relatives. We formalize these statements in the following definition.

Definition 1.3.1 *Let R_n^* be the maximal growth rate achievable in a set of portfolio sequences B when given future knowledge of $X_1, X_2, \ldots, X_n$. We call a portfolio sequence $\left\{\widehat{b}_i\right\}_{i=1}^n$ determined independently of future knowledge a* ***universal portfolio*** *(or simply* ***universal****) with respect to B if its corresponding growth rate $\widehat{R}_n$ is such that,*

$$\lim_{n\to\infty} \sup_{\{X_i\}_{i=1}^n} \left(R_n^* - \widehat{R}_n\right) = \lim_{n\to\infty} \sup_{\{X_i\}_{i=1}^n} \frac{1}{n} \log \frac{W_n^*}{\widehat{W}_n} \leq 0.$$

Here $W_n^* = W_0 \prod_{i=1}^n b_i^{*\prime} X_i$ and $\widehat{W}_n = W_0 \prod_{i=1}^n \widehat{b}_i^\prime X_i$ are the wealths of the best sequence in B chosen in hindsight and the universal portfolio respectively. Additionally, we refer to B as the ***target class*** *of the universal portfolio, R_n^* as the* ***target growth rate****, and $W_n^* = W_0 \exp\{nR_n^*\}$ as the* ***target wealth.***

We reiterate that a universal portfolio may be (and often has to be) outside its target class. We also reiterate that the convergence of growth rates is not in an almost sure

Year	Wealth
89	\$1
90	\$10.18
91	\$459
92	\$39, 881.83
93	\$1.2 Mil
94	\$39.6 Mil
95	\$678.3 Mil
96	\$21.0 Bil
97	\$518.0 Bil
98	\$22.8 Tril

Table 1.1: Wealth achieved by investing in the best stock each day.

sense. We carry no stochastic assumptions whatsoever on the sequences of wealth relatives so this convergence holds uniformly over all permissible sequences.

As we have hinted to earlier, we can't expect to find universal procedures with respect to every class B. This is would be hoping for too much. An example of a set that we have no reasonable hope of being universal with respect to is the set of all permissible portfolio sequences, i.e.,

$$B = \left\{ \{b_i\}_{i=1}^{n} : \sum_{j=0}^{m} b_{i,j} = 1 \text{ and } b_{i,j} \geq 0 \text{ for } 1 \leq i \leq n \text{ and } 0 \leq j \leq m \right\}.$$

If an investor is given knowledge of future prices and is then allowed to use any portfolio sequence he likes, his best strategy would be to put all of his wealth in the best stock each period. This strategy yields the absolute best growth achievable amongst a group of stocks. The resulting wealth can grow incredibly fast. To emphasize how fast we have experimented with this strategy on a small set data. The data consists of dividend adjusted prices for Wells Fargo (WFC), Boise Cascade (BCC), and Exxon (XON) from January 17, 1989 to January 17, 1998. Table 1.1 shows how quickly a dollar would grow if we placed all of our money in the best stock (or cash) each day. By the end of nine years the dollar would have grown into about \$23 trillion, which rivals the annual GDP of the United States! Clearly we cannot hope to track this phenomenal growth without future knowledge of prices.

The point of this experiment is to point out that their are some target wealths that we are not likely to achieve. However if we reduce the size of the target class to some subset of all allowable portfolio sequences we can sometimes get universal portfolios that have excellent asymptotic properties. In particular, the seminal work of Cover [5] introduced the first universal portfolio on the target class of constant rebalanced portfolios. As this portfolio is central to the rest of this thesis we now define it and examine it's properties in some detail.

1.3.2 Cover's Universal Portfolio wrt Constant Rebalanced Portfolios

The concept of a universal portfolio was first put forth by Cover [5] in the context of the constant rebalanced portfolio target class. Recall that a constant rebalanced portfolio (CRP) is a sequence of portfolios for which the wealth in each stock is re-proportioned at the start of evenly spaced periods to bring allocations back to some initial value. Thus if we use portfolio b for the first period we would buy and sell appropriate amounts of stock at the end of each period to insure that we are invested in portfolio b at the start of the next period. The set of all such portfolios is referred to as B_{CRP} and is represented by the m-dimensional simplex

$$B_{\mathsf{CRP}} = \left\{ b \in \mathbf{R}_{+}^{m+1} : \sum_{j=0}^{m} b_j = 1 \text{ and } b_j \geq 0 \text{ for } 0 \leq j \leq m \right\}.$$

The wealth achieved by CRP b is expressed as,

$$W_n(b) = \prod_{i=1}^{n} b' X_i. \tag{1.2}$$

Given future knowledge of prices, we define the *best constant rebalanced portfolio* to be the portfolio b^* that maximizes wealth $W_n(b)$. We define W_n^* to be the maximal wealth achieved by b^*, and label its associated optimal growth rate, $R_n^* = \frac{1}{n} \log W_n^*/W_o$. In terms of constructing a universal portfolio, we view B_{CRP} as the target class, W_n^* as the target wealth, and R_n^* as the target rate.

Before discussing how Cover constructs a universal portfolio for this target class, we first examine some important properties of the target wealth. First and foremost we must realize that W_n^* is bigger than the wealth achieved by putting all money in the best stock. This is so because W_n^* is maximized over all CRP's and a portfolio that continually puts all assets into one stock is a valid CRP. Therefore W_n^* is necessarily bigger than the wealth associated with the best stock.

Another immediate consequence is that W_n^* outperforms various stock indices like the Dow Jones Industrial Average (DJIA). The DJIA is a weighted arithmetic average of stock price with weights $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that $\lambda_j \geq 0$ and $\sum_{j=1}^{m} \lambda_j = 1$. If e_j is a constant rebalanced portfolio putting weight 1 on stock j then DJIA $= \sum_{j=1}^{m} \lambda_j W_n(e_j) \leq \sum_{j=1}^{m} \lambda_j W_n^* = W_n^*$. Similarly W_n^* will also outperform geometric average indices like the Value Line Index. Note that ValueLine $= \left(\prod_{j=1}^{m} (W_n(e_j)/W_0)\right)^{1/m} \leq \left(\prod_{j=1}^{m} (W_n^*/W_0)\right)^{1/m} = W_n^*/W_0$.

Given these properties, many investors would be happy to achieve W_n^*. However we can't use the best CRP b^* that achieves it because we don't have knowledge of future prices. A universal portfolio would circumvent this issue and let us achieve W_n^* to first order in the exponent.

Surprisingly such universal portfolios exists. The first such procedure was presented by Cover [5] and begins by splitting wealth evenly between all stocks and cash, i.e. $\widehat{b}_1 = \left(\widehat{b}_{1,0}, \widehat{b}_{1,1}, \dots, \widehat{b}_{1,m}\right) = \left(\frac{1}{m+1}, \dots, \frac{1}{m+1}\right)$. For each successive time period, wealth is reallocated according to the formula,

$$\widehat{b}_i = \frac{\int_{B_{\text{CRP}}} b W_{i-1}(b) d\mu(b)}{\int_{B_{\text{CRP}}} W_{i-1}(b) d\mu(b)}. \tag{1.3}$$

Here $\mu(b)$ is the uniform measure on B_{CRP} normalized such that $\mu(B_{\text{CRP}}) = 1$. An important property of the portfolio is that its wealth $\widehat{W}_n$ is the uniform average of the wealths achieved by CRP's. In other words,

$$\widehat{W}_n = \int_{B_{\text{CRP}}} W_n(b) d\mu(b).$$

Cover and Ordentlich [6] take advantage of this property to prove that,

$$\frac{W_n^*}{\widehat{W}_n} \leq (n+1)^m. \tag{1.4}$$

Thus the wealth of the procedure comes within a polynomial bound of W_n^*. Such a polynomial bound is certainly sufficient for universality since,

$$\lim_{n\to\infty} \sup_{\{X_i\}_{i=1}^n} \frac{1}{n} \log \frac{W_n^*}{\widehat{W}_n} = \lim_{n\to\infty} \frac{m}{n} \log(n+1) = 0.$$

The portfolio sequence, $\widehat{b}_i$, has a nice interpretation as an implementation of a strategy where we split the initial wealth over a continuum of investment managers, each of whom uses a unique CRP b. To clarify this, suppose that at the start of each period, each manager invests according to a unique CRP. At the end of the period each manager should have a wealth proportional to $W_i(b)$. If originally we had split our initial wealth in proportion to $\mu(b)$, we would expect to have a collective wealth at time n of $\widehat{W}_n = \int_{B_{\text{CRP}}} W_n(b) d\mu(b)$, but this can be rigorously shown to be the wealth achieved by $\widehat{b}_i$. So in essence the use of $\widehat{b}_i$ is equivalent to distributing initial wealth among a continuum of investment managers.

1.3.3 Cover-Ordentlich Universal Portfolio with Side Information

Cover and Ordentlich [6] have extended the above procedure to allow investors to use various sources of *side information* (i.e. past prices, economic indicators, expert opinion, etc.) to update their portfolios. It is assumed that the side information can be modelled according to a variable s taking values in a finite set. The quantity $s_i \in \mathcal{S} = \{1, \ldots, k\}$ denotes the state of side information at the beginning of period i. Depending on the state s_i, the investor is given the freedom to invest in a CRP corresponding to that state. In this way, investment is tailored to the currently available information.

As an example of how side information might be used, technical traders sometimes use "break-out" signals to determine when a stock might be a good buy opportunity. One rule might be to give greater weight to a stock j that has outperformed other stocks over the last 10 trading periods. In this case we would set $s_i = j$, and invest in the CRP corresponding to this state of side information. Thus the investor has the flexibility to change his investment preferences according to what he considers to be the best performing stock of late.

Extending our previous notation, we use $b(1), b(2), \ldots, b(k)$ to denote the portfolios used under each state of side information. Thus the portfolio $b(s_i)$ represents the portfolio used in period i. The wealth achieved using portfolio mapping $b(\cdot)$ is given by,

$$W_n(b(\cdot), s^n) = W_0 \prod_{i=1}^{n} b'(s_i) X_i.$$

Given the sequence of wealth relatives $X_1, X_2, \ldots, X_n$, and side information $s^n = (s_1, \ldots s_n)$ we can determine the portfolio mapping achieving the best possible wealth. This is denoted by $b^*(\cdot)$ where,

$$b^*(\cdot) = \arg\max_{b(\cdot)} W_0 \prod_{i=1}^{n} b'(s_i) X_i.$$

Similarly we denote $W_n^*(s^n)$ to be the corresponding maximal wealth.

In order to construct a universal portfolio with respect to the target wealth $W_n^*(s^n)$, Cover and Ordentlich suggest an analog of (1.3). For each period i, the suggested universal portfolio is computed via formula,

$$\widehat{b}_i(s) = \frac{\int_{B_{\mathrm{CRP}}} b S_{i-1}(b|s) d\mu(b)}{\int_{B_{\mathrm{CRP}}} S_{i-1}(b|s) d\mu(b)},$$

where $S_i(b|s)$ is the wealth relative generated by constant rebalanced portfolio b along the subsequence $\{j \leq i : s_j = s\}$, and is given by,

$$S_i(b|s) = \prod_{j \leq i : s_j = s} b'(s) X_j.$$

The resulting wealth of this procedure $\widehat{W}_n = \prod_{i=1}^{n} \widehat{b}_i'(s_i) X_i$ can be shown to come within a polynomial factor of the target wealth $W_n^*(s^n)$. In particular, if $\mu(b)$ is taken to be the uniform measure over B_{CRP} it is shown that,

$$\frac{W_n^*(s^n)}{\widehat{W}_n} \leq (n+1)^{km}. \tag{1.5}$$

It is interesting to compare the above bound with bound (1.4) obtained for the single CPR target class. We see that there is a trade-off involved when using side information. Increasing the number of states k is good in that it increases the target wealth $W_n^*(s^n)$. However it is also bad in that for every state of side information considered, we pay a further price of $(n+1)^m$ in our bound. Thus we are discouraged from considering too many states.

1.3.4 Other Results Pertaining to Universal Portfolios

In addition to Cover's results there are some other works pertaining to universal portfolios that should be mentioned before proceeding further. Helmbold, et al. [8] have recently suggested another universal portfolio which is also universal with respect to the class of CRP's. The primary advantage of this procedure is that it is easier to compute than Cover's, both in terms of time and memory. The drawbacks are that it necessitates absolute bounds on wealth relatives and the wealth of the procedure fails to come within a polynomial bound of W_n^*. We examine this procedure in greater detail in a future section.

Another result that is related to universality is given by Algoet [1]. Recall that Algoet and Cover [2] have derived a portfolio strategy that achieves optimal growth if wealth relatives are generated according to some *known* stationary ergodic process. Algoet [1] has addressed the problem of achieving this optimal growth when the specific stationary ergodic process underlying the wealth relatives is *unknown*. We have discussed previously that if the distribution is known, the growth optimal strategy is characterized by,

$$b_i^*(X_{i-1}, \ldots, X_1) = \arg\max_b E_{P_i} \log(b' X_i)$$

where the expectation is taken with respect to the conditional distribution,

$$P_i = P(dx_i | X_{i-1}, \ldots, X_1).$$

The asymptotic growth rate of this optimal procedure is given by,

$$R^* = \lim_{n\to\infty} E_P \log\left(X_1' b(X_0, X_{-1}, \ldots, X_{-n})\right).$$

where P is the unconditional joint distribution of the process. If P and the conditional distributions P_i are unknown Algoet's approach is to use portfolios $\widehat{b}_i$ that are computed using estimates $\widehat{P}_i$ of the conditional distributions P_i. Specifically,

$$\widehat{b}_i (X_{i-1}, \ldots, X_1) = \arg\max_b E_{\widehat{P}_i} \log \left(b' X_i\right).$$

If the sequence of estimates $\widehat{P}_i$ is such that $\widehat{P}_i \longrightarrow P_i$ weakly almost surely in i, then Algoet shows, under a constraint that keeps wealth relatives bounded away from 0, that the asymptotic growth rate achieved by $\widehat{b}_i$ equals R^* almost surely.

This result is intimately related to the non-stochastic universality results put forth by Cover. In both cases we have some optimal growth rate that we would like to achieve. In turn, the optimal growth rate is computed on the basis of knowledge which we don't have. In Cover's case, the knowledge would be the future outcome of prices. In Algoet's case it would be the knowledge of a specific distribution. In both frameworks, a portfolio sequence is constructed independent of this knowledge that asymptotically achieves the optimal growth rate.

We reiterate that we choose to reserve the term "universal portfolio" for procedures achieving optimal growth rates over arbitrary sequences, independent of stochastic assumptions. Algoet's procedure is in many ways a stochastic analog of the universal portfolio concept. However, because of its stochastic nature we hesitate to refer to it specifically as a universal portfolio.

In a similar vein, Jamshidian [10] has examined growth optimality in continuous time under the assumption that prices behave according to some unknown Gaussian Process. The problem of characterizing the growth optimal procedure in continuous time for a known Gaussian market has been previously solved by Merton [16], [17]. Using the growth rate of Merton's optimal procedure as a target, Jamshidian proposes a portfolio strategy, independent of the specific form of the Gaussian Process, that matches this optimal growth asymptotically with almost sure probability. The proposed procedure is essentially an extension to continuous time of Cover's universal portfolio in discrete time.

1.4 Layout of the Thesis

The goal of this thesis is to extend and refine some of the concepts of universal portfolios as introduced by Cover [5]. Our main goal is to consider target classes that make use of past price information or other available side information, s, in determining portfolio allocations through certain parametric forms of dependence on s. Unlike constant rebalanced portfolios, the proportion of wealth allocated by these portfolios changes from period to period depending on the state of s. As we have seen above, Cover and Ordentlich

[6] have already developed a universal procedure in the case where s takes on a finite number of values. In addition to the finite state case, our more general development encompasses forms of dependence where s is now a variable taking values in a continuum. It is with these new forms of dependence that we find linearly parametrized classes of portfolios that lead us to computationally convenient universal procedures.

The first chapter considers the existence of universal portfolios for parameterized classes in a discrete time setting. Using a simple generalization of the Cover universal portfolio, we construct portfolios that are proven to come within a polynomial factor of the target wealth (i.e. the wealth of the best portfolio in the target class chosen with hindsight) and are hence universal with respect to the target class.

A drawback of these discrete time universal procedures is that they are computationally intensive to compute. For the case of Cover's procedure calculation is contingent upon computing the integral $\int_{B^+} bW(b)d\mu(b)$. For most choices of measure, no closed form solution exists so we must resort to numerical integration. In general these computations are of exponential order in the number of stocks m. In the case that μ is taken to be the Dirichlet$(1/2, \ldots, 1/2)$ distribution it is possible to evaluate the integral directly, but the resulting closed form calculations needed for each period are still of exponential order in the number of stocks, i.e. n^m. Thus, even for a relatively small group of stocks the number of calculations can quickly become prohibitively large. For this reason, it is desirable to find settings for which "nice" closed form solutions exists. Solutions requiring computations that are only polynomial in m would make the corresponding universal procedures much more accessible to strategies involving larger numbers of stocks.

A search for easily computable universal portfolios motivates an extension of results to continuous time. This is the subject of the next chapter of the thesis. Jamshidian [10] has previously developed extensions of the Cover portfolio to the case of continuous time Gaussian markets. We too develop continuous time analogs but this time in a purely non-stochastic framework. The only assumption made is that log price paths exhibit regularity in fluctuation. Namely, the paths must exhibit positive and finite quadratic variation. Although we choose to emphasize the lack of stochastic assumptions, it is worth noting that our results would also hold almost surely for log prices governed by an underlying diffusion process.

Using only properties of log-price paths it is possible to derive expressions of the wealth of a continuously updated constant rebalanced portfolio that agree with those derived by Jamshidian [10], Merton [16], [17], and Larson [11]. The result is an exponential quadratic in b, implying that wealth as a function of b is a Gaussian-shaped curve. At this point one can borrow an idea from Bayesian analysis, namely that of the normal conjugate prior, and produce a weighted portfolio with easily computable components. We go on to prove that this continuously update portfolio is universal with respect to the entire set of constant rebalanced portfolios including those that allow for short selling.

This universal portfolio is extendable to the larger framework where we allow the use of side information. This allows for the development of easily computable universal portfolios for a variety of continuous time target classes.

Though ease of computation of the universal portfolio and its associated wealth at any point in time is certainly desirable, the problem remains of implementing these computations and associated trades on a continuous basis. In order to reconcile a universal portfolio that is easily calculable with one that is tradable we examine a discrete time portfolio in the last chapter of the thesis that is a direct analog of the continuous time procedure. In the end we find that if we trade this easily computable portfolio with an ever increasing frequency, it remains universal with respect to the continuous time target class. In the final section of the thesis, we examine the performance of this portfolio using stock data from the NYSE. In our exhibit we find that our new portfolio outperforms the best stock over the examined time period.

Chapter 2

Universality in Discrete Time

The goal of this chapter is to develop discrete time universal portfolios for parameterized target classes possibly exhibiting continuous form dependence on side information. We begin with some definitions and then discuss specific examples of parameterized target classes for which we might find universal procedures. We proceed to work with an analog of Cover's universal portfolio for these target classes and then show that it is universal. Finally we discuss some of the computational problems related to these portfolios and how they might be overcome.

2.1 Preliminaries

We begin with a discrete time examination of investment in a market of m stocks and cash. For each investment period $i = 1, 2, \ldots, n$ the investor allocates his wealth according to some portfolio of cash and stocks specified by the vector $b_i = (b_{i,0}, \widetilde{b}_i) = (b_{i,0}, b_{i,1}, \ldots, b_{i,m})$. Here, $b_{i,0}$ represents the proportion of wealth put in cash, and the vector $\widetilde{b}_i = (b_{i,1}, \ldots, b_{i,m})$ represents the proportions of wealth put in each stock. To insure the portfolios are self financing we require that each portfolio be in the set $\{b \in \mathbf{R}^{m+1} : \sum_{j=0}^{m} b_j = 1\}$. For the time being we assume that both short selling and buying on margin are forbidden and hence require that $b_{i,j} \geq 0$ for each i, j. We refer to the set of all such portfolios as B^+ and write,

$$B^+ = \{b \in \mathbf{R}^{m+1} : \sum_{j=0}^{m} b_j = 1,\ b_j \geq 0 \text{ for all } j\}.$$

For each investment period we denote the price of stock j at the end of period i by $P_{i,j}$. Given these prices we define the *wealth relative*, $X_{i,j} = P_{i,j}/P_{i-1,j}$, to be the ratio of the price of stock j at the end of period i to that at the beginning of period i. Collectively we write the vector of wealth relatives as,

$$X_i = (1, X_{i1}, \ldots, X_{im}),$$

with the understanding that the first component, the wealth relative of cash, is always 1. We think of wealth relatives as the factor by which wealth increases over a period. Thus if a stock has a wealth relative of 1.03, it implies the stock increased 3% over the period.

Given these definitions we see that an investor starting with initial wealth W_0 and investing in the sequence of portfolios $b_1, \ldots, b_n$, yields a wealth after n periods of,

$$W_n = W_0 \prod_{i=1}^{n} b_i' X_i.$$

Recall that our present goal is to define target classes (i.e. sets of portfolio sequences) for which we will find universal portfolios. Instead of considering the whole set of arbitrary portfolio sequences, we wish to restrict attention to various classes of finitely parameterized portfolio sequences. We also wish to consider the use of available side information (i.e. past prices, economic indicators, expert opinion, etc.) by sequences in these classes to influence the allocation of wealth at the start of each period. If such information is used, we will assume that it can be summarized at the start of period i through the state variable s_i which takes values in some domain S. We assume little about the domain and allow it to be discrete or continuous and of arbitrary dimension.

To define a class of parameterized portfolio sequences, we define a parameter space $\Theta \subseteq \mathbf{R}^d$, and a portfolio map $b : \Theta \times S \rightarrow B^+$ that for each period i sets the portfolio $b_i \equiv b(\theta, s_i)$. We think of the class as a set of functions $\{b(\theta, \cdot) : \theta \in \Theta\}$. Each function in the class, or equivalently each $\theta \in \Theta$, defines a distinct sequence of portfolios whose allocations are determined at the start of period i through $b(\theta, s_i)$.

This framework is sufficiently general to model a large number of portfolio sequence families and provides a plethora of potential target classes for universal procedures. In particular, some of the target classes considered by other authors now become special cases. To better illustrate the role of classes, the use of side information, etc., we now give some specific examples.

2.2 Some Parameterized Classes of Portfolio Sequences

2.2.1 Constant Rebalanced Portfolios

We have already encountered the class of constant rebalanced portfolios in the introduction. Recall that constant rebalanced portfolios are portfolio strategies that keep the same portfolio for each trading period. Thus, if b_1 is used in period 1, we buy and sell enough stock so that at the start of each subsequent period i, $b_i = b_1$.

One possible parameterization of this class is given by the parameter space $\Theta = \{\theta \in \mathbf{R}^m : \sum_{j=1}^m \theta_j \leq 1, \theta_j \geq 0\}$ and mapping $b_i \equiv b(\theta) = (1 - \sum_{j=1}^m \theta_j, \theta_1, \ldots, \theta_m)$. Since constant rebalanced portfolios do not use side information, the state variable s_i is left undefined.

The class of constant rebalanced portfolios and its associated target wealth achieved by the best constant rebalanced portfolio have some nice properties. In the case of an i.i.d. market, Merton [17] and Breiman [4] have shown that the best constant rebalanced portfolio chosen in hindsight is asymptotically growth optimal with probability 1. Without stochastic assumptions, it is always the case that the best constant rebalanced portfolio outperforms the best buy and hold strategy (and hence the best stock) for any price sequence. Thus any universal procedure associated with this class is also guaranteed to beat or match the wealth of the best stock in the long run.

2.2.2 A Simple Class using Side Information

We previously discussed how Cover and Ordentlich have incorporated side information into a universal procedure when the number of states of s is finite. Now we see how we can reformulate their approach in our current setting. Consider the investor who uses k different portfolios $b(1), \ldots, b(k)$ depending on the current state of side information $s_i \in \{1, \ldots, k\}$. As an example, the side information in this case could be a sliding scale of an experts "bullishness" about the market. The investor sets $s_i = 1$ if the expert is very bearish at the start of period i and $s_i = k$ if he is very bullish. Values between 1 and k would then reflect an opinion between the two extremes. Depending on the outcome, s_i, the investor uses portfolio $b_i = b(s_i)$. In this manner, the investor is able to tailor his investment strategy according to expert opinion.

Each distinct set of portfolios $b(1), \ldots, b(k)$, corresponds to a distinct strategy. Thus to parameterize this class we need to parameterize the set of all such collections of k portfolios and then select the portfolio used depending on the value of s_i. A way to formulate this would be to use the parameter space,

$$\Theta = \{\theta \in \mathbf{R}^{k(m+1)} : \sum_{j=0}^{m} \theta_j \leq 1, \sum_{j=m+1}^{2(m+1)} \theta_j \leq 1, \ldots, \sum_{j=(k-1)(m+1)}^{k(m+1)-1} \theta_j \leq 1,\ \theta_j \geq 0\},$$

and mapping $b(\theta, s_i) = (\theta_{(s_i-1)(m+1)}, \ldots, \theta_{s_i(m+1)-1})$.

2.2.3 A Portfolio Class with Continuous Form Dependence on Side Information

The previous class is formulated to handle side information in a discrete form. In cases where side information is more readily expressed in terms of a continuous s, the above procedure can be adapted by discretizing s into sufficiently small partitions. However, as we have discussed in the introduction, there is a trade-off involved in increasing the number of states. For each new state considered there is a corresponding increase in the dimensionality of Θ and of the polynomial bound determining the maximum deviation of the universal portfolio wealth from target wealth. Perhaps more importantly, increasing the number of states exponentially increases the computational complexity of the corresponding universal portfolio. This is bad. The alternative is to form classes that depend on s in a continuous fashion. This can eliminate the need for excessively large parameter spaces and allows continuous form side information to be used in a natural way.

A particular type of side information well suited to continuous use is past prices. In a market where prices trend, it might be reasonable to use past prices to determine how wealth should be allocated. In particular, for such a market it makes sense to put more wealth in stocks that have shown stronger performance in recent periods. One strategy might be to invest in the most recent wealth relative vector. For example, at the start of period i, we could invest according to $X_{i-1}/\sum_{j=0}^{m} X_{i-1,j}$. However, an investor should feel uncomfortable using only past price information to set his allocations in this manner. Firstly, the basis of the strategy is contrary to the weak form of the efficient market hypothesis which suggests that such trending should not exist. Secondly, the suggested portfolio is contra to the buy low, sell high paradigm of constant rebalanced portfolios which tends to work well in trendless (i.e. i.i.d.) markets. Thus, to guard against overreliance on trending the investor may want to split his money between $X_{i-1}/\sum_{j=0}^{m} X_{i-1,j}$ and a constant rebalanced portfolio.

Thus consider the class of portfolio sequence for which, before investment, the investor fixes a constant rebalanced portfolio and a fraction of wealth to put between the constant rebalanced portfolio and most recent wealth relative vector. In this case, our side information is the normalized wealth relative vector $s_i = X_{i-1}/\sum_{j=0}^{m} X_{i-1,j}$. Our parameters θ are used to set the constant rebalanced portfolio and the fraction of wealth. Hence a possible parameterization of the class would be given by parameter space $\Theta = \{\theta \in \mathbf{R}^{m+1} : \sum_{j=1}^{m} \theta_j \leq 1, \theta_j \geq 0\ \forall j, \theta_{m+1} \leq 1\}$ and mapping $b_i \equiv b(\theta, s_i) = (1-\theta_{m+1})s_i + \theta_{m+1}(1-\sum_{j=1}^{m}\theta_j, \theta_1, \ldots, \theta_m)$.

2.3 Universal Portfolios for Parameterized Classes

2.3.1 Goal Statement

We now endeavor to find a universal portfolio with respect to a parameterized target class defined by a closed parameter space $\Theta \subseteq \mathbf{R}^d$, side information domain S, and portfolio mapping $b(\theta, s)$. Given a sequence of wealth relative vectors $X_1, X_2, \ldots, X_n$ and side information states $s_1, \ldots, s_n$ the wealth achieved by the portfolio sequence indexed by θ up to time n is given by,

$$W_n(\theta) = W_0 \prod_{i=1}^{n} b'(\theta, s_i) X_i.$$

If knowledge of stock prices were known up to n periods into the future, it would be possible to determine the optimal wealth attainable within the target class up to time n. As before we refer to this optimal wealth as the *target wealth* and denote it by, $W_n^* \equiv \max_{\theta \in \Theta} W_n(\theta)$. The target wealth might be achieved by more than one sequence. However, regardless of the uniqueness of these sequences, we use the symbol θ_n^* to refer to an index of an optimizing sequence as of time n, and write $W_n^* = W_n(\theta_n^*)$.

Since $W_n(\theta_n^*)$ is a product of factors it tends to grow exponentially with n. For a portfolio sequence indexed by θ we again define the *growth rate* to time n as $R_n(\theta) = \frac{1}{n}\log(W_n(\theta)/W_0)$. Hence,

$$W_n(\theta) = W_0 \exp\{n\, R_n(\theta)\}. \tag{2.1}$$

From (2.1) we see that we can essentially achieve $W_n(\theta_n^*)$ if we invest in a portfolio sequence $\widehat{b}_i$ with wealth $\widehat{W}_n$ achieving a growth rate $\widehat{R}_n$ such that $\lim_{n\to\infty}(\widehat{R}_n - R_n(\theta_n^*)) = 0$. Recall from definition 1.3.1 that we will call such a $\widehat{b}_i$ a *universal portfolio* if it does not depend on knowing prices in advance and achieves the optimal growth rate asymptotically and uniformly over all sequences, i.e.,

$$\lim_{n\to\infty} \sup_{\{X_i\}_{i=1}^n} (\widehat{R}_n - R_n(\theta_n^*)) = \lim_{n\to\infty} \sup_{\{X_i\}_{i=1}^n} \frac{1}{n} \log \frac{W_n(\theta_n^*)}{\widehat{W}_n} \leq 0.$$

It is also correct to say that $\widehat{b}_i$ is universal with respect to a target class if it achieves the target wealth $W_n(\theta^*)$ up to first order in the exponent. In the next section we show one method of constructing such a procedure for a general parameterized target class.

2.3.2 Constructing a Universal Portfolio

In order to construct a universal procedure for a parameterized target class, we are motivated to begin with the procedure proposed by Cover [5] for the target class of constant rebalanced portfolios. Recall that for a constant rebalanced portfolio $b \in B^+$ and associated wealth to time i, $W_n(b) = W_0 \prod_{i=1}^n b' X_i$, Cover has shown that the portfolio sequence,

$$\widehat{b_i} = \frac{\int_{B^+} b W_{i-1}(b) d\mu(b)}{\int_{B^+} W_{i-1}(b) d\mu(b)},$$

is universal with respect to target wealth $W_n^* = \max_{b \in B^+} W(b)$. (Here, $\mu(b)$ is the uniform measure on B^+).

With the hope that this universal property can be preserved, we are motivated to generalize this sequence to a target class with parameter space $\Theta \subseteq \mathbf{R}^d$, side information space S, and portfolio mapping $b(\theta, s)$. Defining π to be a measure on the Borel σ-field of Θ of total measure 1, the obvious generalization of this portfolio is,

$$\widehat{b_i} = \frac{\int_\Theta b(\theta, s_i) W_{i-1}(\theta) d\pi(\theta)}{\int_\Theta W_{i-1}(\theta) d\pi(\theta)}. \tag{2.2}$$

Thus the portfolio constitutes a weighted average of portfolios in the target class weighted according to how well these portfolios have done in the past. We denote the wealth achieved by $\widehat{b_i}$ up to time n by $\widehat{W_n} = W_0 \prod_{i=1}^n \widehat{b}_i' X_i$. The following lemma shows that this wealth is nicely expressible as the π-weighted average wealth of the target class.

Lemma 2.3.1 *The wealth, $\widehat{W_n}$, achieved by portfolio sequence (2.2) can be written as,*

$$\widehat{W_n} = \int_\Theta W_n(\theta) d\pi(\theta).$$

Proof. The proof follows from a telescoping product. Note that,

$$\begin{aligned}
\widehat{W_n} &= W_0 \prod_{i=1}^n \widehat{b}_i X_i \\
&= W_0 \prod_{i=1}^n \frac{\int_\Theta b'(\theta, s_i) W_{i-1}(\theta) d\pi(\theta)}{\int_\Theta W_{i-1}(\theta) d\pi(\theta)} X_i \\
&= W_0 \prod_{i=1}^n \frac{\int_\Theta b'(\theta, s_i) X_i W_{i-1}(\theta) d\pi(\theta)}{\int_\Theta W_{i-1}(\theta) d\pi(\theta)} \\
&= W_0 \prod_{i=1}^n \frac{\int_\Theta W_i(\theta) d\pi(\theta)}{\int_\Theta W_{i-1}(\theta) d\pi(\theta)} \\
&= W_0 \frac{\int_\Theta W_n(\theta) d\pi(\theta)}{\int_\Theta W_0 d\pi(\theta)} \\
&= \int_\Theta W_n(\theta) d\pi(\theta)
\end{aligned}$$

■

The portfolio sequence, $\widehat{b}_i$, can be interpreted as an implementation of a strategy where we split the initial wealth over a continuum of investment managers, each of whom uses a unique portfolio sequence in the target class indexed by θ. To clarify we suppose that at the start of each period, each manager invests according to his θ in the portfolio $b(\theta, s_i)$. At the end of the period each manager has wealth proportional to $W_i(\theta)$. If originally we had split our initial wealth in proportion to $\pi(\theta)$, we would expect to have a collective wealth at time n of $\widehat{W}_n = \int_\Theta W_n(\theta)d\pi(\theta)$, but as was shown in the above lemma this is the same wealth achieved by $\widehat{b}_i$. So in essence the use of $\widehat{b}_i$ is equivalent to distributing initial wealth among a continuum of investment managers.

We now intend to show that $\widehat{b}_i$ of equation (2.2) is indeed universal respect to its associated target class under certain conditions. To this point, the behavior of target classes and price sequences has been left unspecified. However in order to get a workable environment for universality we need to place conditions on each of these. Depending on the strength of the result we wish to show, we will henceforth use two sets of assumptions. The first, weaker set is as follows.

Weak Investment Assumptions

W1 Wealth relatives are bounded in the sense that there exists some constant $L_X > 0$ such that $1/L_X \le X_{i,j} \le L_X$ for all $1 \le i \le n$ and $0 \le j \le m$.

W2 Θ is a convex, compact subset of $\mathbf{R}^d$ having positive Lebesgue measure with respect to $\mathbf{R}^d$.

W3 The mapping $b(\theta, s)$ is Lipschitz in that there exists a constant $L_b > 0$ independent of θ and s such that $||b(\theta_0, s) - b(\theta_1, s)|| \le L_b||\theta_0 - \theta_1||$ for all $\theta_0, \theta_1 \in \Theta$ and $s \in S$.

We also employ a stronger set of assumptions which keeps conditions W1 and W2 but which replaces continuity of $b(\theta, s)$ with a differentiability condition. We also add a further condition that a subset of the maximizing parameters θ_n^* must eventually lie within the interior of the parameter space.

Strong Investment Assumptions

S1 W1 and W2 hold.

S2 The mapping $b(\theta, s)$ is twice differentiable in θ for all s and has bounded second order partials in the sense that there exist a constant $M > 0$ such that for every $1 \le i \le n$,

$s \in S$, and $\theta = (\phi_1, \dots, \phi_d) \in \Theta$ each component, $b_j(\theta, s)$, of the portfolio mapping satisfies

$$\left| \frac{\partial^2 b_j(\theta, s)}{\partial \phi_q \partial \phi_r} \right| < M$$

for every $q, r \in \{1, \dots, d\}$.

S3 There exists integer N such that for $n \geq N$, a maximizing parameter,

$$\theta_n^* \in \arg\max_{\theta \in \Theta} W_n(\theta),$$

lies in the interior of Θ.

With these sets of assumptions we now present the following theorem that proves the universality of our proposed universal procedure.

Theorem 2.3.2 *Suppose the Weak Investment Assumptions hold. Suppose also that $\pi(\theta)$, a measure on the Borel σ-field of $\Theta \subset \mathbf{R}^d$ with $\pi(\Theta) = 1$, is absolutely continuous with respect to Lebesgue measure and has corresponding Radon-Nikodym derivative $f(\theta)$ which is bounded below on Θ by some $\delta > 0$. Then the portfolio sequence $\widehat{b}_i$ given in (2.2) is universal with respect to the target class $(\Theta, S, b(\theta, s))$ and target wealth $W_n(\theta_n^*) = \max_{\theta \in \Theta} W_n(\theta)$ in the sense that,*

$$\frac{W_n(\theta_n^*)}{\widehat{W}_n} = O\left(n^d\right).$$

Furthermore if the Strong Investment Assumptions hold, the order of the bound is improved to,

$$\frac{W_n(\theta_n^*)}{\widehat{W}_n} = O\left(n^{d/2}\right).$$

The order bounds imply that $\widehat{b}_i$ is universal since,

$$\lim_{n \to \infty} \sup_{\{X_i\}_{i=1}^n} \frac{1}{n} \log \frac{W_n(\theta_n^*)}{\widehat{W}_n} = \lim_{n \to \infty} O\left(\frac{\log n}{n}\right) = 0.$$

Thus $\widehat{b}_i$ achieves the best wealth chosen in hindsight up to first order in the exponent. We now proceed to prove the theorem.

Proof. We start with the Weak Investment Assumptions and prove the $O\left(n^d\right)$ result. We wish to examine the behavior of wealths around the neighborhood of an optimal parameter θ_n^*. For this reason we define the neighborhood $\Theta_n^* = \{\theta \in \Theta : \|\theta - \theta_n^*\| \leq 1/n\}$. In contrast to θ_n^*, which achieves maximal wealth, we define the parameter $\theta_n^\dagger$ to be a member of the subset of parameters that achieves minimal wealth among elements of Θ_n^*. Hence, we select $\theta_n^\dagger \in \arg\min_{\Theta_n^*} W_n(\theta)$ among a set of minimizers. At each period i, we define the difference between two portfolios attaining the target wealth and the minimal wealth by, $\Delta b_i(\theta_n^*, \theta_n^\dagger, s_i) \equiv b_i(\theta_n^*, s_i) - b_i(\theta_n^\dagger, s_i)$. Now we note the following inequality,

$$
\begin{aligned}
\Delta b_i'(\theta_n^*, \theta_n^\dagger, s_i)X_i &= \left[b_i(\theta_n^*, s_i) - b_i(\theta_n^\dagger, s_i)\right]' X_i \\
&\leq L_b\|\theta_n^* - \theta_n^\dagger\|\|X_i\| \text{ by W3} \\
&\leq \sqrt{m}L_X L_b\|\theta_n^* - \theta_n^\dagger\| \text{ by virtue of } \max_{i,j} X_{i,j} < L_X \\
&\leq \frac{\sqrt{m}L_X L_b}{n} \text{ by definition of } \Theta_n^*.
\end{aligned}
$$

Now we use Lemma 2.3.1 to note that,

$$
\begin{aligned}
\frac{\widehat{W_n}}{W_n^*} &= \int_\Theta \frac{W_n(\theta)}{W_n^*} d\pi(\theta) \\
&\geq \int_{\Theta_n^*} \frac{W_n(\theta)}{W_n^*} d\pi(\theta) \\
&= \int_{\Theta_n^*} \frac{\prod_{i=1}^n b'(\theta, s_i)X_i}{\prod_{i=1}^n b'(\theta_n^*, s_i)X_i} d\pi(\theta) \\
&\geq \frac{\prod_{i=1}^n b'(\theta_n^\dagger, s_i)X_i}{\prod_{i=1}^n b'(\theta_n^*, s_i)X_i} \pi(\Theta_n^*) \\
&= \prod_{i=1}^n \left[1 + \frac{\Delta b_i'(\theta_n^*, \theta_n^\dagger, s_i)X_i}{b'(\theta_n^\dagger, s_i)X_i}\right]^{-1} \pi(\Theta_n^*),
\end{aligned}
$$

which, after using the bound on $\Delta b_i'(\theta_n^*, \theta_n^\dagger, s_i)X_i$, is lower bounded by,

$$
\geq \prod_{i=1}^n \left[1 + \frac{\sqrt{m}L_X L_b}{n b'(\theta_n^\dagger, s_i)X_i}\right]^{-1} \pi(\Theta_n^*),
$$

which in turn by using $b'(\theta_n^\dagger, s_i)X_i \geq 1/L_X$ is lower bounded by,

$$
\begin{aligned}
&\geq \prod_{i=1}^n \left[1 + \frac{\sqrt{m}L_X^2 L_b}{n}\right]^{-1} \pi(\Theta_n^*) \\
&= \left(1 + \frac{\sqrt{m}L_X^2 L_b}{n}\right)^{-n} \pi(\Theta_n^*) \\
&\geq e^{-\sqrt{m}L_X^2 L_b}\pi(\Theta_n^*).
\end{aligned}
$$

The rest of the proof for the $O\left(n^d\right)$ case hinges on bounding $\pi(\Theta_n^*)$. We prove in Lemma 5.2.1 of the appendix that there exists a constant $R > 0$ such that $\pi(\Theta_n^*) \geq Rn^{-d}$. Thus,

$$\frac{\widehat{W}_n}{W_n^*} \geq e^{-\sqrt{m}L_X^2 L_b} Rn^{-d}. \tag{2.3}$$

Inverting the ratio we conclude that $\frac{W_n^*}{\widehat{W}_n}$ is $O(n^d)$.

Now we proceed with the proof of the $O(n^{d/2})$ case under the Strong Investment Assumptions. As before, we consider the behavior of wealth around the neighborhood of a maximizer θ_n^*. However, instead of shrinking the neighborhood on order $1/n$ we decrease shrinkage to order $1/\sqrt{n}$. Thus we define neighborhood $\Theta^*_{\sqrt{n}} \equiv \{\theta \in \Theta : \|\theta - \theta_n^*\| \leq \frac{1}{\sqrt{n}}\}$. Define

$$l_n(\theta) \equiv \frac{1}{n}\log(W_n(\theta)/W_0) = \frac{1}{n}\sum_{i=1}^{n}\log b^T(\theta, s_i)x_i.$$

We wish to bound $l_n(\theta)$ in the neighborhood of $\Theta^*_{\sqrt{n}}$. By Taylor's Theorem,

$$l_n(\theta) = l_n(\theta_n^*) + \sum_{q=1}^{d}(\phi_q - \phi_q^*)\frac{\partial l_n(\theta_n^*)}{\partial \phi_q} + \frac{1}{2}\sum_{q,r=1}^{d}(\phi_q - \phi_q^*)(\phi_r - \phi_r^*)\frac{\partial^2 l_n(\theta')}{\partial\phi_q\partial\phi_r},$$

for some $\theta' \in \Theta$ where ϕ_q and ϕ_q^* are the qth co-ordinates of θ and θ_n^* respectively. The first thing to note is that for every q,

$$\frac{\partial l_n(\theta_n^*)}{\partial \phi_q} = 0.$$

This is a necessary condition for θ_n^* to be the maximizing strategy if it is in the interior of Θ. We now turn our attention to bounding $\frac{\partial^2 l_n(\theta)}{\partial\phi_q\partial\phi_r}$ in the neighborhood $\Theta^*_{\sqrt{n}}$ which in turn entails the first and second order partials of $b_j(\theta, s_i)$. If the second order partials of $b_j(\theta, s_i)$ are bounded uniformly by M then, due to the compactness of Θ, we can make M sufficiently large to bound the first partials as well. The absolute value of the second partial of $l_n(\theta)$ is bounded as follows,

$$\begin{aligned}
\left|\frac{\partial^2 l_n(\theta)}{\partial\phi_q\partial\phi_r}\right| &= \left|\frac{1}{n}\sum_{i=1}^{n}\left(\left[b^T(\theta, s_i)x_i\right]^{-1}\sum_{j=1}^{m}x_{ij}\frac{\partial^2 b_j(\theta, s_i)}{\partial\phi_q\partial\phi_r}\right)\right. \\
&\quad \left. -\frac{1}{n}\sum_{i=1}^{n}\left(\left[b^T(\theta, s_i)x_i\right]^{-2}\sum_{j,k=1}^{m}x_{ij}x_{ik}\frac{\partial b_k(\theta, s_i)}{\partial\phi_q}\frac{\partial b_j(\theta, s_i)}{\partial\phi_r}\right)\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}L_X\sum_{j=1}^{m}L_X M + \frac{1}{n}\sum_{i=1}^{n}L_X^2\sum_{j,k=1}^{m}L_X^2 M^2 \\
&\leq mL_X^2 M + m^2 L_X^4 M^2.
\end{aligned}$$

For convenience we set $K = mL_X^2 M + m^2 L_X^4 M^2$. Now note the following,

$$\begin{aligned}
\frac{\widehat{W}_n}{W_n^*} &= \int_\Theta \frac{W_n(\theta)}{W_n^*} d\pi(\theta) \\
&\geq \int_{\Theta^*_{\sqrt{n}}} \frac{W_n(\theta)}{W_n^*} d\pi(\theta) \\
&= \int_{\Theta^*_{\sqrt{n}}} \exp\left\{ n\frac{1}{n}\log(W_n(\theta)/W_0) - n\frac{1}{n}\log(W_n^*/W_0)\right\} d\pi(\theta) \\
&= \int_{\Theta^*_{\sqrt{n}}} \exp\left\{ nl_n(\theta_n^*) + \frac{n}{2}\sum_{q,r=1}^{d}(\phi_q - \phi_q^*)(\phi_r - \phi_r^*)\frac{\partial^2 l_n(\theta')}{\partial\phi_q\partial\phi_r} \right. \\
&\qquad \left. -nl_n(\theta_n^*)\right\} d\pi(\theta) \\
&= \int_{\Theta^*_{\sqrt{n}}} \exp\{\frac{n}{2}\sum_{q,r=1}^{d}(\phi_q - \phi_q^*)(\phi_r - \phi_r^*)\frac{\partial^2 l_n(\theta')}{\partial\phi_q\partial\phi_r}\} d\pi(\theta) \\
&\geq \int_{\Theta^*_{\sqrt{n}}} \exp\{-\frac{n}{2}\sum_{q,r=1}^{d}\|\phi_q - \phi_q^*\|\|\phi_r - \phi_r^*\|K\} d\pi(\theta) \\
&\geq \exp\{-\frac{d^2}{2}K\}\pi(\Theta^*_{\sqrt{n}}) \quad (\text{as } \|\phi_q - \phi_q^*\| \leq \frac{1}{\sqrt{n}})
\end{aligned}$$

As before, we show in Lemma 5.2.1 of the appendix that there exists a constant $R > 0$ such that $\pi(\Theta^*_{\sqrt{n}}) \geq Rn^{-d/2}$. Thus we have,

$$\frac{\widehat{W}_n}{W_n^*} \geq \exp\{-\frac{d^2}{2}K\}Rn^{-d/2} \tag{2.4}$$

Inverting the ratio we conclude that $\frac{W_n^*}{\widehat{W}_n}$ is $O(n^{d/2})$. ■

The significance of Theorem 2.3.2 is that we can use (2.2) to construct a universal portfolio for any target class that satisfies our investment assumptions. It can be easily verified that the sample target classes in sections 2.2.1 - 2.2.3 satisfy at least the Weak Investment Assumptions so in each case $\widehat{b}_i$ results in a universal procedure.

It is also interesting to note that the resulting procedures for the target classes in sections 2.2.1 and 2.2.2 are equivalent to those suggested in Cover and Ordentlich [6]. However, the constants derived here for the polynomial bounds, (2.3) and (2.4), are necessarily cruder than those previously published since we have assumed less about the nature of the target classes. Obviously it is possible to improve these bounds if we can take advantage of the specific nature of $b_j(\theta, s_i)$, Θ and π.

2.4 Computational Issues

The universality of $\widehat{b}_i$ is certainly desirable. The problem remains of computing it. As was noted in the introduction, the calculation of $\widehat{b}_i$ requires numerical methods due to the absence of a closed form solution to $\int_\Theta b(\theta, s) W_n(\theta) d\pi(\theta)$. Although numerical integration offers an approximation, the order of calculation grows exponentially with the dimensionality of Θ. For classes using very modest numbers of stocks (and hence dimensionality of Θ), the calculations involved quickly become prohibitive.

In the context of the constant rebalanced portfolio (CRP) class, there has been an effort to develop other universal procedures which exhibit better computational properties. Helmbold, et. al. [8] have developed a universal procedure for which computations increase only linearly with the number of stock. The resulting portfolio is constructed in part by using an iterative update based on the portfolio b_i^* maximizing the objective function,

$$F(b_i) = \eta \log (b_i' X_{i-1}) - D(b_i || b_{i-1}).$$

Here, $\eta > 0$ is a parameter called the learning rate and $D(u||v) \equiv \sum_{j=1}^{m} u_j \log \frac{u_j}{v_j}$ is the relative entropy distance that acts as penalty term. The actual universal procedure is calculating by taking the maximal b^* and shrinking it slightly towards the vector $\left(\frac{1}{m+1}, \dots, \frac{1}{m+1}\right)$. Both the magnitude of this shrinking and the value of the learning rate decrease as a function of n. Since the maximization of the objective is shown to be linear in m and constant in n the algorithm is very efficient.

In contrast to these methods, we draw inspiration from a different direction in order to get computationally convenient portfolios for parameterized target classes. Rather than circumvent the calculation of $\int_\Theta b(\theta, s) W_n(\theta) d\pi(\theta)$ we search for cases where the calculation is direct through the existence of convenient closed form solutions. While these solutions remain elusive in discrete time, we will show that solutions exist for linearly parameterized classes traded in continuous time that are of computational order d^2 where d is the dimensionality of our parameter space Θ.

Chapter 3

Universality in Continuous Time

In this chapter we extend and modify some of the concepts of the previous chapter to the case of continuous time trading. As before we work in the absence of stochastic assumptions and derive all our results using only assumed properties of the price paths themselves. We begin by examining the behavior of constant rebalanced portfolios under continuous trading and derive a closed form expression for their wealth. We then go on to consider how side information might be incorporated at discrete time intervals. This results in parameterized portfolio classes that act as continuously traded constant rebalanced portfolios over the discrete time periods. We then construct a universal procedure with respect to these classes by taking a limit of discrete procedures. The result is a direct continuous time analog of the discrete universal portfolio (2.2). After proving universality for the general case, we examine classes that are linear in their parameterization and find that under the correct choice of Gaussian measure, these classes yield easily computable universal procedures. The last theorem of the chapter derives a worst case bound for the wealth of these portfolios.

3.1 Continuously Traded Constant Rebalanced Portfolios

Let us consider investment in a constant rebalanced portfolio (CRP) among m stocks and cash. As before we denote this portfolio by the vector $b = (b_0, \widetilde{b}) = (b_0, b_1, \dots, b_m)$ where the vector $\widetilde{b} = (b_1, \dots, b_m)$ holds the proportions of wealth put in each stock and b_0 denotes the proportion of wealth put in cash. To insure that the portfolio is self financing we require that $\sum_{j=0}^{m} b_j = 1$. However, unlike the discrete case, we no longer assume that the b_j are non-negative. In other words we now allow for short selling and purchase on margin. For convenience we henceforth refer to the set of all such portfolios as B, and write

$$B = \left\{ b \in R^m : \sum_{j=0}^{m} b_j = 1 \right\}.$$

Consider investing in such a portfolio when we rebalance n times over an arbitrary time horizon $T \in \mathbf{R}^+$. Viewing stock prices as a realization of some continuous function of time we write $P_{t,j}$ to denote the price of stock j at time $t \in [0, T]$. As before, we suppose that we start with some initial wealth W_0. For any time $t \leq T$, we write the wealth produced by b up to the last rebalancing by time t as,

$$\begin{aligned} W_t^{(n)}(b) &= W_0 \prod_{k=1}^{\lfloor nt/T \rfloor} \left(b_0 + \sum_{j=1}^{m} b_j \frac{P_{kT/n,j}}{P_{(k-1)T/n,j}} \right) \\ &= W_0 \prod_{k=1}^{\lfloor nt/T \rfloor} \left(1 + \sum_{j=1}^{m} b_j \left(\frac{P_{kT/n,j}}{P_{(k-1)T/n,j}} - 1 \right) \right). \end{aligned} \tag{3.1}$$

We would like to know what happens to $W_t^{(n)}(b)$ as n goes to ∞ (i.e. when we are rebalancing the portfolio continuously). We seek an expression for the limiting wealth $W_t(b) = \lim_{n\to\infty} W_t^{(n)}(b)$ for an arbitrary price path. To derive this expression it will be convenient at times to work with the log price path $Z_t \equiv (\log P_{t,1}, \dots, \log P_{t,m})$. For any realization of this path we define the empirical log-drift up to time t as,

$$\mu_t = (\mu_{t,1}, \dots, \mu_{t,m}) = (Z_{t,1} - Z_{0,1}, \dots, Z_{t,m} - Z_{0,m}).$$

Similarly we also define the sequence of empirical covariation matrices $K_t^{(n)}$ with entries,

$$\begin{aligned} K_{t,i,j}^{(n)} = \sum_{k=1}^{\lfloor nt/T \rfloor} & \left(Z_{(k/n)T,i} - Z_{((k-1)/n)T,i} \right) \\ \times & \left(Z_{(k/n)T,j} - Z_{((k-1)/n)T,j} \right). \end{aligned}$$

In order to derive a limiting wealth $W_t(b)$ we will need to assume that price paths are not too wild in their fluctuation. The primary property they need to exhibit is finite quadratic variation. We would also like the empirical covariation matrices to converge to some positive definite limit. Henceforth we require price paths to meet the following assumptions:

Minimal Path Assumptions

P1 There exists a constant $L_P > 0$ dependent on the path and time horizon T such that

$$\left(1+\frac{L_P(1+\log n)}{\sqrt{n}}\right)^{-1} \le \frac{P_{kT/n,j}}{P_{(k-1)T/n,j}} \le 1+\frac{L_P(1+\log n)}{\sqrt{n}}$$

for all n, $k \in \{1,\ldots,n\}$, and $j \in \{1,\ldots,m\}$.

P2 At each instant t, there exists a positive definite matrix K_t such that $\lim_{n\to\infty} K_t^{(n)} = K_t$ element-wise. Moreover, $K_t - K_s$ is positive definite for all $s < t$.

Given these assumptions we now derive an expression for the wealth achieved by continuously trading a constant rebalanced portfolio. It should come as no surprise that the expression we derive is in agreement with those previously published by Merton [17], Larson [11], and Jamshidian [10]. However, unlike these results which are proven using an underlying diffusion process for P_t, we choose to stay away from stochastic assumptions and instead use only path properties. This non-stochastic setting is consistent with our goal of developing universal procedures that have growth optimal properties independent of distributional assumptions.

Theorem 3.1.1 *Let $b = (b_0, \tilde{b}) \in B$ be a constant rebalanced portfolio (with short selling and leveraging permitted). If the Minimal Path Assumptions hold, the wealth at time t from trading b continuously is,*

$$W_t(b) = W_0 \exp\left\{\mu_t'\tilde{b} + \frac{1}{2}\sum_{j=1}^{m} K_{t,j,j}b_j - \frac{1}{2}\tilde{b}'K_t\tilde{b}\right\}. \tag{3.2}$$

Proof. See appendix. ∎

Expression (3.2) is an example of how analysis is sometimes simpler when working in continuous time. Recall that in discrete time the universal portfolios we considered required computation of an integral of the form $\int bW_t^{(n)}(b)d\pi(b)$. Computing this integral is often problematic when using the discrete wealth $W_t^{(n)}(b)$ as given in (3.1). However upon going to continuous time, $W_t^{(n)}(b)$ becomes $W_t(b)$ and we can exploit the form of (3.2) to make computation of the integral much easier. Realizing that (3.2) is an exponential quadratic in b,we can choose μ to be Gaussian and use the idea of a normal conjugate prior to argue that $W_t(b)d\pi(b)$ is a non-normalized Gaussian measure. In this case the integral $\int bW_t^{(n)}(b)d\pi(b)$ is essentially equivalent to a normal expectation. This idea will later be key to our development of easily computable universal procedures for linearly parameterized target classes.

Before proceeding to the next section we pause first to prove a corollary to Theorem 3.1.1 identifying the best constant rebalanced portfolio chosen in hindsight. As before we denote the hindsight optimal wealth as $W_t^* = \max_{b \in B} W_t(b)$. Since $W_t(b)$ is an exponential quadratic in b, the best CRP is simply the portfolio b_t^* that maximizes this quadratic. We identify b^* as follows,

Corollary 3.1.2 *Under the conditions of Theorem 3.1.1 , the optimal wealth, W_t^*, is achieved by portfolio $b_t^* = \left(b_{t,0}^*, \widetilde{b}_t^*\right)$ with,*

$$\widetilde{b}_t^* = K_t^{-1}\left(\mu_t + \frac{1}{2}\text{diag}(K_t)\right). \tag{3.3}$$

As always, $\widetilde{b}_t^* = (b_{t,1}^*, \dots, b_{t,m}^*)$ denotes the optimal proportions in stock and $b_{t,0}^* = 1 - \sum_{j=1}^{m} b_{t,j}^*$ denotes the optimal proportion in cash. Furthermore, optimal wealth can be expressed as,

$$W_t^* = W_0 \exp\{\frac{1}{2}\widetilde{b}_t^{*\prime} K_t \widetilde{b}_t^*\} = W_0 \exp\left\{\frac{1}{2}(\mu_t + \frac{1}{2}diag(K_t))' K_t (\mu_t + \frac{1}{2}diag(K_t))\right\}.$$

Proof. It is easy to verify that for b_t^* given in (3.3),

$$-\frac{1}{2}(\widetilde{b} - \widetilde{b}_t^*)' K_t (\widetilde{b} - \widetilde{b}_t^*) + \frac{1}{2}\widetilde{b}_t^{*\prime} K_t \widetilde{b}_t^* = \mu_t' \widetilde{b} + \frac{1}{2}\sum_{j=1}^{m} K_{t,j,j} b_j - \frac{1}{2}\widetilde{b}' K_t \widetilde{b}$$

But the right hand side is just the exponent of the wealth expression (3.2). Thus it follows that $W_t(b)$ is maximized when $-\frac{1}{2}(\widetilde{b} - \widetilde{b}_t^*)' K_t (\widetilde{b} - \widetilde{b}_t^*)$ is maximized. Since K_t is positive definite the preceding quadratic can be at most 0 so it follows that the quadratic is maximized when $\widetilde{b} - \widetilde{b}_t^* = 0$ or $\widetilde{b} = \widetilde{b}_t^*$. Given the expression (3.3) of $\widetilde{b}_t^*$, the corresponding expressions for optimal wealth follow immediately. ■

3.2 Continuous Time Portfolio Classes using Side Information

3.2.1 Specification of the Classes

We now use the continuously traded constant rebalanced portfolios examined in the previous section to create new classes of continuously traded portfolios having dependence on side information (e.g. past prices, expert opinion, earnings reports, etc.). As before, we assume that the side information is represented by a variable s taking values in domain S. Even though we will be trading in continuous time we choose to apply this side information only at the start of T discrete time periods indexed by $\tau \in \{1, \dots, T\}$. For ease of interpretation, we might think of these periods as weeks or months or any interval for which it would be reasonable to collect and use the side information. The idea will be to use the side information to select a constant rebalanced portfolio at the start of each period and then continuously trade that portfolio for the rest of the period.

To be more rigorous we define a parameter space $\Theta \subseteq \mathbf{R}^d$ and portfolio mapping $b : \Theta \times S \to B$. At the start of period 1 we use the available side information s_1 and portfolio mapping $b(\theta, s)$ to set the constant rebalanced portfolio $b_1 = b(\theta, s_1)$. We then take b_1 and trade it continuously over time period $t \in (0, 1]$. At the start of the next period we take side information s_2 and set the constant rebalanced portfolio $b_2 = b(\theta, s_2)$ which we then trade continuously over the time period $t \in (1, 2]$. This process is repeated T times until we reach our invest horizon at time $t = T$.

As in the previous chapter we think of a particular *class* as a triplet of parameter space, side information domain, and portfolio map, i.e. $(\Theta, S, b(\theta, s))$. Each member of a class is indexed by a parameter $\theta \in \Theta$ and represents a sequence of T constant rebalanced portfolios traded continuously over T time periods.

We use $W_t(\theta)$ to denote the wealth achieved by $b(\theta, s)$ as of time t. Since $b(\theta, s_\tau)$ is just a succession of CRP's we can easily apply Theorem 3.1.1 to obtain an expression of wealth. To apply the theorem to the present case we must first define analogs of the empirical log-drift μ_t and covariance matrix K_t that measure drifts and covariances from the beginning of the most recent time period as of time t. For this reason we define the drift vector,

$$\mu_t^\dagger \equiv (\mu_{t,1}^\dagger, \ldots, \mu_{t,m}^\dagger) = (Z_{t,1} - Z_{\lceil t \rceil - 1, 1}, \ldots, Z_{t,m} - Z_{\lceil t \rceil - 1, m})$$

and covariance matrix $K_t^\dagger \equiv \lim_{n \to \infty} K_t^{\dagger(n)}$ where $K_t^{\dagger(n)}$ has entries,

$$\begin{aligned} K_{t,i,j}^{\dagger(n)} = \sum_{k=1}^{\lfloor n(t - \lceil t \rceil + 1) \rfloor} & \left(Z_{\lceil t \rceil - 1 + k/n, i} - Z_{\lceil t \rceil - 1 + (k-1)/n, i}\right) \\ \times & \left(Z_{\lceil t \rceil - 1 + k/n, j} - Z_{\lceil t \rceil - 1 + (k-1)/n, j}\right). \end{aligned}$$

Setting the index set $I(t) = \{1, \ldots, \lceil t \rceil - 1, t\}$ and successively applying Theorem 3.1.1 to each period prior to and including time t,we conclude that for any time $0 \le t \le T$,

$$\begin{aligned} W_t(\theta) = & W_0 \exp\left\{ \sum_{\tau \in I(t)} \left(\mu_\tau^\dagger\right)' \tilde{b}(\theta, s_{\lceil \tau \rceil}) + \frac{1}{2} \sum_{\tau \in I(t)} \sum_{j=1}^{m} K_{\tau,j,j}^\dagger b_j(\theta, s_{\lceil \tau \rceil}) \right. \\ & \left. -\frac{1}{2} \sum_{\tau \in I(t)} \tilde{b}'(\theta, s_{\lceil \tau \rceil}) K_\tau^\dagger \tilde{b}(\theta, s_{\lceil \tau \rceil}). \right. \end{aligned} \tag{3.4}$$

(Recall here that $\widetilde{b}(\theta, s)$ is the vector of stock proportions associated with the portfolio mapping $b(\theta, s)$). More simply, at the end of T time periods can we write,

$$W_T(\theta) = W_0 \exp\left\{\sum_{\tau=1}^{T}\left(\mu_\tau^{\dagger}\right)'\widetilde{b}(\theta, s_\tau) + \frac{1}{2}\sum_{\tau=1}^{T}\sum_{j=1}^{m} K_{\tau,j,j}^{\dagger} b_j(\theta, s_t) - \frac{1}{2}\sum_{\tau=1}^{T}\widetilde{b}'(\theta, s_\tau) K_\tau^{\dagger}\widetilde{b}(\theta, s_\tau)\right\}. \tag{3.5}$$

As before, we denote the optimal wealth (or target wealth) within the class of strategies by,

$$W_t^* \equiv \max_{\theta \in \Theta} W_t(\theta),$$

and refer to a parameter that achieves this maximum by θ_t^*.

3.2.2 Example

We now present an example of a class using side information to help clarify this framework. Recall from Corollary 3.1.2 of Section 3.1 that the optimal constant rebalanced portfolio for period τ given K_τ and μ_τ is $b_\tau^* = \left(b_{\tau,0}^*, \widetilde{b}_\tau^*\right)$ where,

$$\widetilde{b}_\tau^* = K_\tau^{-1}\left(\mu_\tau + \frac{1}{2}\text{diag}(K_\tau)\right).$$

and $b_{\tau,0}^* = 1 - \sum_{j=1}^{m} b_{\tau,j}^*$. If the behavior of the market is changing slowly over time, this portfolio should be a good predictor of the best constant rebalanced portfolio over the next period. In this case it makes sense to invest in $\widetilde{b}_{\tau-1}^*$ at the beginning of period τ. To be safe though we might also want to diversify in a constant rebalanced portfolio remaining constant over all periods.

We could represent this type of strategy by setting the parameters,

$$\theta = (\theta_1, \ldots, \theta_m, \theta_{m+1}) \in \mathbf{R}^{m+1},$$

side information vector $s_\tau = \widetilde{b}_{\tau-1}^*$, and portfolio map $b(\theta, s) = \left(b_0(\theta, s), \widetilde{b}(\theta, s)\right)$, where

$$\widetilde{b}(\theta, s) = \theta_{m+1}s + (\theta_1, \ldots, \theta_m),$$

and $b_0(\theta, s) = 1 - \sum_{j=1}^{m} b_j(\theta, s)$. Thus, θ_{m+1} is used as a weighting parameter for $\widetilde{b}_{\tau-1}^*$ and the vector $(\theta_1, \ldots, \theta_m)$ represents the stock weightings of the constant rebalanced portfolio. At the start of time period τ we would begin to continuously rebalance the portfolio,

$$b\left(\theta, \widetilde{b}_{\tau-1}^*\right) = \left(1 - \sum_{j=1}^{m}\left(\theta_{m+1}\widetilde{b}_{\tau-1,j}^* + \theta_j\right), \theta_{m+1}\widetilde{b}_{\tau-1,1}^* + \theta_1, \ldots, \theta_{m+1}\widetilde{b}_{\tau-1,m}^* + \theta_m\right).$$

The fact that θ is allowed to take any value in $\mathbf{R}^{m+1}$ implies that the components of this portfolio can take any value in $\mathbf{R}$. This might seem confusing at first because it is customary to think of proportions of wealth taking values between 0 and 1. However there are still reasonable interpretations for values lying outside this range. Consider the first component representing cash, i.e. $1 - \sum_{j=1}^{m} \left(\theta_{m+1}\widetilde{b}^{*}_{\tau-1,j} + \theta_j\right)$. A negative value for this quantity indicates that we are in a net *leveraged* position or borrowing money to buy greater amounts of stock. A value greater than 1 means that we are short selling stock and holding it as cash. As for the components representing the stocks themselves, a negative value indicates that we are short selling the stock (i.e. selling borrowed stock that we promise to replace later) and a value greater than 1 means we are borrowing money (or short selling other assets) to buy more of the stock than we could otherwise afford.

3.3 A Continuous Time Universal Procedure for a General Class

Our goal now is to find a universal procedure for the parameterized and continuously traded target classes of the previous section. As before in the discrete time case we will need to restrict attention to classes that satisfy some minimal properties. In particular, we will need to assume that parameter spaces are closed and convex and that portfolio mappings are continuous. These properties are rigorously stated in the following minimal class conditions.

Minimal Class Conditions

1. $\Theta \subseteq \mathbf{R}^d$ is a closed and convex.
2. The mapping $b(\theta, s)$ is Lipschitz in the sense that there exists a constant $L_b > 0$ independent of θ and s such that $\|b(\theta_0, s_i) - b(\theta_1, s_i)\| \leq L_b\|\theta_0 - \theta_1\|$ for all $\theta_0, \theta_1 \in \Theta$ and $s \in S$.

By definition, a universal portfolio for our present type of target class $(\Theta, S, b(\theta, s))$ must be a non-anticipating portfolio $\widehat{b}_t$ that generates wealth $\widehat{W}_t$ matching the hindsight optimal wealth $W_t^* \equiv \max_{\theta\in\Theta} W_t(\theta)$ to first order in the exponent. The universal portfolio (2.2) already accomplishes this for the discrete case. Thus an intuitive way to produce a universal portfolio for the present case might be to adapt (2.2) to continuous time. The adaptation would simply involve trading (2.2) on a finer and finer time scale until in limit we would be continuously trading the portfolio,

$$\widehat{b}_t = \frac{\int_\Theta b(\theta, s_{\lfloor t\rfloor+1}) W_t(\theta) d\pi(\theta)}{\int_\Theta W_t(\theta) d\pi(\theta)}, \tag{3.6}$$

at each time instance t. The concept of what it means to trade $\widehat{b}_t$ at each time instance can be unsettling. However, we stay rigorous by defining $\widehat{b}_t$ as a limit of discrete procedures and by proving results pertaining to $\widehat{b}_t$ by way of this discrete limit representation. This will be key in proving the following lemma which gives an expression for the wealth achieved by trading $\widehat{b}_t$ continuously.

Lemma 3.3.1 *Suppose that the minimal path and class conditions are satisfied. Let $\widehat{W}_t$ be the wealth achieved by trading $\widehat{b}_t$ continuously over time interval* $[0, t]$. *Then,*

$$\widehat{W}_t = \int_\Theta W_t(\theta) d\pi(\theta). \tag{3.7}$$

Proof. The proof consists of dominated convergence arguments showing that the proposed $\widehat{b}_t$ (3.6) is the limit of discrete procedures. In parallel we also show that wealth (3.7) is also the limit of wealth of these same procedures.

We start by considering a discrete analog of the continuous trading schemes we have considered thus far. We begin by defining a portfolio map $b(\theta, s)$ that is used to set CRP's at the beginnings of T time periods. In turn these CRP's are rebalanced n times over their respective periods. Thus $b(\theta, s_1)$ represents the first CRP rebalanced n times over period (0,1], $b(\theta, s_2)$ represents the CRP rebalanced over period (1,2], etc. The wealth achieved by this process up to the most recent rebalancing by time $t \leq T$ is,

$$W_t^{(n)}(\theta) = W_0 \prod_{k=1}^{\lfloor nt\rfloor} \left(b_0(\theta, s_{\lceil k/n\rceil}) + \sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n\rceil}) \frac{P_{k/n,j}}{P_{(k-1)/n,j}} \right).$$

Because components of $b(\theta, s)$ can take negative values it is possible that this wealth might be negative. We would like to avoid this if possible. We can circumvent this issue by restricting attention to a compact subset of the parameter space. For this reason we define for $\lambda > 0$ and $\theta_0 \in \Theta$ the compact set $\Theta_\lambda = \{\theta \in \Theta : \|\theta - \theta_0\| \leq \lambda\}$. Given the continuity of $b(\theta, s)$ we can always find an upper bound depending on the sequence of side information outcomes $s_1, \ldots, s_T$ that bounds the magnitude of the components of $b(\theta, s)$ on this set. Given this bound we can always find a sufficiently large N such that for $n \geq N$ all factors of the wealth relative $\prod_{k=1}^{\lfloor nt\rfloor} b(\theta, s_{\lceil k/n\rceil}) X_i$ are positive and hence the wealth $W_t^{(n)}(\theta)$, is positive.

For a given price path and side information sequence, we would now like to show that the wealth $W_t^{(n)}(\theta)$ is uniformly bounded over all $n \geq N$, and $\theta \in \Theta_\lambda$. This will be needed later in the proof when we apply dominated convergence arguments. Note that since $\sum_{j=0}^{m} b_j(\theta, s) = 1$,

$$W_t^{(n)}(\theta) = W_0 \prod_{k=1}^{\lfloor nt \rfloor} \left(1 + \sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n \rceil}) \left(\frac{P_{k/n,j}}{P_{(k-1)/n,j}} - 1 \right) \right).$$

Setting,

$$y_{kj}^{(n)} \equiv \left(P_{k/n,j}/P_{(k-1)/n,j} - 1 \right)$$

and using the Taylor expansion $y = \log(1+y) + y^2/(2(1+c)^2)$ on each $y_{kj}^{(n)}$we write,

$$W_t^{(n)}(\theta) = W_0 \prod_{k=1}^{\lfloor nt \rfloor} \left(1 + \sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n \rceil}) \left(\log(1 + y_{kj}^{(n)}) + \frac{\left(y_{kj}^{(n)}\right)^2}{2(1+c_{kj})^2} \right) \right).$$

for c_{kj} between 0 and $y_{kj}^{(n)}$. Recall that for $n \geq N$, $W_t^{(n)}(\theta)$ is positive. Therefore for such n each factor in the above product is also positive. For any product $\prod_{i=1}^{n} x_i$ with positive factors the following inequality holds, $\prod_{i=1}^{n} x_i = \exp\left\{\sum_{i=1}^{n} \log x_i\right\} \leq \exp\left\{\sum_{i=1}^{n} (x_i - 1)\right\}$. Applying this to the above equation yields,

$$\begin{aligned} W_t^{(n)}(\theta) &\leq W_0 \exp\left\{ \sum_{k=1}^{\lfloor nt \rfloor} \sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n \rceil}) \left(\log(1 + y_{kj}^{(n)}) + \frac{\left(y_{kj}^{(n)}\right)^2}{2(1+c_{kj})^2} \right) \right\} \\ &= W_0 \exp\left\{ \sum_{k=1}^{\lfloor nt \rfloor} \sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n \rceil}) \log(1 + y_{kj}^{(n)}) \right. \\ &\quad \left. + \sum_{k=1}^{\lfloor nt \rfloor} \sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n \rceil}) \frac{\left(y_{kj}^{(n)}\right)^2}{2(1+c_{kj})^2} \right\}. \end{aligned} \tag{3.8}$$

We bound each term of the RHS of (3.8) separately. Note that,

$$\begin{aligned} \sum_{k=1}^{\lfloor nt \rfloor} \sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n \rceil}) \log(1 + y_{kj}^{(n)}) &= \sum_{j=1}^{m} \sum_{k=1}^{\lfloor nt \rfloor} b_j(\theta, s_{\lceil k/n \rceil}) \log(1 + y_{kj}^{(n)}) \\ &= \sum_{j=1}^{m} \left(\sum_{k=1}^{n} b_j(\theta, s_1) \log(1 + y_{kj}^{(n)}) \right. \\ &\quad + \sum_{k=n+1}^{2n} b_j(\theta, s_2) \log(1 + y_{kj}^{(n)}) + \cdots \\ &\quad \left. + \sum_{k=n\lfloor t \rfloor + 1}^{\lfloor nt \rfloor} b_j(\theta, s_{\lceil t \rceil}) \log(1 + y_{kj}^{(n)}) \right). \end{aligned}$$

By definition of $y_{kj}^{(n)}$, we can write $\sum_{k=1}^{n} \log(1+y_{kj}^{(n)}) = \log(P_{1,j}/P_{0,j})$. By the minimal path assumptions $|\log(P_{1,j}/P_{0,j})| \leq L_P \log(1+T)/\sqrt{T}$. Similarly,

$$\sum_{k=n+1}^{2n} \log(1+y_{kj}^{(n)}) \leq L_P \log(1+T)/\sqrt{T}, \ldots, \sum_{k=n\lfloor t\rfloor+1}^{\lfloor nt\rfloor} \log(1+y_{kj}^{(n)}) \leq L_P \log(1+T)/\sqrt{T},$$

so,

$$\begin{aligned} & \sum_{k=1}^{\lfloor nt\rfloor}\sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n\rceil}) \log(1+y_{kj}^{(n)}) \qquad (3.9) \\ = \; & L_P \log(1+T)/\sqrt{T} \sum_{j=1}^{m} \left(b_j(\theta,s_1) + b_j(\theta,s_2) + \ldots + b_j(\theta, s_{\lceil t\rceil})\right). \\ \leq \; & mL_P \log(1+T)/\sqrt{T} \max_{1\leq j\leq m, 1\leq\tau\leq\lceil t\rceil, \theta\in\Theta_\lambda} b_j(\theta, s_\tau). \end{aligned}$$

From the minimal class assumptions $b_j(\theta, s_\tau)$ is a continuous mapping on Θ_λ for any choice of j and τ. Since Θ_λ is compact, it follows that a finite maximum is attained by $b_j(\theta, s_\tau)$ over $\theta \in \Theta_\lambda$ for each j and τ. Hence, $\max_{j,\tau,\theta} b_j(\theta, s_\tau)$ is finitely bounded by some constant. Thus the LHS of (3.9) can be bounded independently of θ and n.

Now consider the second term of (3.8),

$$\sum_{k=1}^{\lfloor nt\rfloor}\sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n\rceil}) \frac{\left(y_{kj}^{(n)}\right)^2}{2(1+c_{kj})^2}. \qquad (3.10)$$

Recall from the minimal path assumptions that,

$$\left|\log(1+y_{kj}^{(n)})\right| \leq \frac{L_P(1+\log n)}{\sqrt{n}}, \qquad (3.11)$$

and,

$$\left|y_{kj}^{(n)}\right| = \left|P_{k/n,j}/P_{(k-1)/n,j} - 1\right| \leq \frac{L_P(1+\log n)}{\sqrt{n}}. \qquad (3.12)$$

Also since c_{kj} is some number between 0 and $y_{kj}^{(n)}$ it follows that,

$$2(1+c_{kj})^2 \geq 2(1 - \frac{L_P(1+\log n)}{\sqrt{n}})^2.$$

However since this bound for $2(1+c_{kj})^2$ is ultimately converging to 2 we can increase N as necessary to insure that for some positive constant C,

$$2(1+c_{kj})^2 \geq C, \quad \text{for } n \geq N. \qquad (3.13)$$

Using (3.13) we bound the second term of (3.8) by,

$$\sum_{k=1}^{\lfloor nt\rfloor}\sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n\rceil})\frac{\left(y_{kj}^{(n)}\right)^2}{2(1+c_{kj})^2} \leq \max_{1\leq j\leq m, 1\leq \tau\leq \lceil t\rceil, \theta\in\Theta_\lambda} b_j(\theta, s_\tau)\sum_{j=1}^{m}\sum_{k=1}^{\lfloor nt\rfloor}\frac{\left(y_{kj}^{(n)}\right)^2}{C}.$$

We have already argued that there is some constant C_2 such that,

$$\max_{1\leq j\leq m, 1\leq \tau\leq \lceil t\rceil, \theta\in\Theta_\lambda} b_j(\theta, s_\tau) \leq C_2$$

so it follows that,

$$\sum_{k=1}^{\lfloor nt\rfloor}\sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n\rceil})\frac{\left(y_{kj}^{(n)}\right)^2}{2(1+c_{kj})^2} \leq C_2\sum_{j=1}^{m}\sum_{k=1}^{\lfloor nt\rfloor}\frac{\left(y_{kj}^{(n)}\right)^2}{C}.$$

Again using the Taylor expansion $y = \log(1+y) + y^2/(2(1+c)^2)$ on each $y_{kj}^{(n)}$ on the RHS of the previous equation we write,

$$\begin{aligned}
&\sum_{k=1}^{\lfloor nt\rfloor}\sum_{j=1}^{m} b_j(\theta, s_{\lceil k/n\rceil})\frac{\left(y_{kj}^{(n)}\right)^2}{2(1+c_{kj})^2} \\
\leq\ & \frac{C_2}{C}\sum_{j=1}^{m}\sum_{k=1}^{\lfloor nt\rfloor}\left(\log\left(1+y_{kj}^{(n)}\right) + \frac{\left(y_{kj}^{(n)}\right)^2}{2(1+c_{kj})^2}\right)^2 \\
=\ & m\frac{C_2}{C}\sum_{j=1}^{m}\left(\sum_{k=1}^{\lfloor nt\rfloor}\log^2\left(1+y_{kj}^{(n)}\right) + \sum_{k=1}^{\lfloor nt\rfloor}\log\left(1+y_{kj}^{(n)}\right)\frac{\left(y_{kj}^{(n)}\right)^2}{2(1+c_{kj})^2}\right. \\
&\left. + \sum_{k=1}^{\lfloor nt\rfloor}\frac{\left(y_{kj}^{(n)}\right)^4}{(2(1+c_{kj})^2)^2}\right).
\end{aligned} \tag{3.14}$$

Now we bound each of the terms in (3.14). First note that,

$$\begin{aligned}
\sum_{k=1}^{\lfloor nt\rfloor}\log^2\left(1+y_{kj}^{(n)}\right) &= \sum_{k=1}^{\lfloor nt\rfloor}\log^2\left(P_{k/n,j}/P_{(k-1)/n,j}\right) \\
&= \sum_{k=1}^{\lfloor nt\rfloor}\left(Z_{k/n,j} - Z_{(k-1)/n,j}\right)^2 \\
&= K_{t,j,j}^{\dagger(n)}.
\end{aligned}$$

But under the minimal path conditions $K_{t,j,j}^{\dagger(n)}$ converges to the finite limit $K_{t,j,j}^{\dagger}$ so it follows that the first term of (3.14) is bounded uniformly over $n \geq N$. As for the second term of

(3.14) note that by using (3.11), (3.12), and (3.13) we can bound this term according to,

$$\begin{aligned}\sum_{k=1}^{\lfloor nt \rfloor} \log\left(1 + y_{kj}^{(n)}\right) \frac{\left(y_{kj}^{(n)}\right)^2}{2(1+c_{kj})^2} &\leq \sum_{k=1}^{\lfloor nt \rfloor} \frac{L_P^3 (1+\log n)^3}{Cn^{3/2}} \\ &\leq t\frac{L_P^3 (1+\log n)^3}{Cn^{1/2}}.\end{aligned}$$

Clearly this latter bound is itself bounded uniformly over $n \geq N$, so we conclude that the second term of (3.14) is bounded uniformly over $n \geq N$. Turning attention to the third term of (3.14) we note that,

$$\begin{aligned}\sum_{k=1}^{\lfloor nt \rfloor} \frac{\left(y_{kj}^{(n)}\right)^4}{(2(1+c_{kj})^2)^2} &\leq \sum_{k=1}^{\lfloor nt \rfloor} \frac{L_P^4 (1+\log n)^4}{C^2 n^2} \\ &\leq t\frac{L_P^4 (1+\log n)^4}{C^2 n}.\end{aligned}$$

Hence we also conclude that the third term of (3.14) is bounded uniformly over $n \geq N$. Since we have succeeded in showing that each term of (3.14) is bounded uniformly over $n \geq N$ it follows from (3.14) that the second term of (3.8) (i.e. (3.10)) is bounded uniformly over $n \geq N$ and $\theta \in \Theta_\lambda$. Since we have now shown that both terms in (3.8) are bounded uniformly over $n \geq N$ and $\theta \in \Theta_\lambda$ we conclude that $W_t^{(n)}(\theta)$ is bounded uniformly over $n \geq N$ and $\theta \in \Theta_\lambda$.

Now consider the sequence of discretely updated portfolios,

$$\widehat{b}_{t,\lambda}^{(n)} = \frac{\int_{\Theta_\lambda} b(\theta, s_{\lfloor t \rfloor + 1}) W_t^{(n)}(\theta) d\pi(\theta)}{\int_{\Theta_\lambda} W_t^{(n)}(\theta) d\pi(\theta)}.$$

and the sequence of wealth relatives,

$$X_k^{(n)} = \left(1, \frac{P_{k/n,1}}{P_{(k-1)/n,1}}, \frac{P_{k/n,2}}{P_{(k-1)/n,2}}, \ldots, \frac{P_{k/n,m}}{P_{(k-1)/n,m}}\right).$$

For initial wealth W_0, the wealth $\widehat{W}_{t,\lambda}^{(n)}$ achieved by $\widehat{b}_{t,\lambda}^{(n)}$ as of the most recent rebalance is,

$$\begin{aligned}
\widehat{W}_{t,\lambda}^{(n)} &= W_0 \prod_{k=1}^{\lfloor nt \rfloor} \widehat{b}_{(k-1)/n,\lambda}^{(n)} X_k^{(n)} \\
&= W_0 \prod_{k=1}^{\lfloor nt \rfloor} \frac{\int_{\Theta_\lambda} b(\theta, s_{\lfloor (k-1)/n \rfloor + 1}) X_k^{(n)} W_{(k-1)/n}^{(n)}(\theta) d\pi(\theta)}{\int_{\Theta_\lambda} W_{(k-1)/n}^{(n)}(\theta) d\pi(\theta)} \\
&= W_0 \prod_{k=1}^{\lfloor nt \rfloor} \frac{\int_{\Theta_\lambda} W_{k/n}^{(n)}(\theta) d\pi(\theta)}{\int_{\Theta_\lambda} W_{(k-1)/n}^{(n)}(\theta) d\pi(\theta)} \\
&= W_0 \frac{\int_{\Theta_\lambda} W_{\lfloor nt \rfloor/n}^{(n)}(\theta) d\pi(\theta)}{\int_{\Theta_\lambda} W_0 d\pi(\theta)} \\
&= W_0 \frac{\int_{\Theta_\lambda} W_t^{(n)}(\theta) d\pi(\theta)}{W_0} \\
&= \int_{\Theta_\lambda} W_t^{(n)}(\theta) d\pi(\theta).
\end{aligned}$$

Now make n large. As n gets larger we rebalance more and more frequently until, at the limit, we are trading in continuous time. Set $\widehat{b}_{t,\lambda} \equiv \lim_{n\to\infty} \widehat{b}_{t,\lambda}^{(n)}$. Since $W_t^{(n)}(\theta)$ is uniformly bounded over $n \geq N$ and $\theta \in \Theta_\lambda$ it holds that $W_t^{(n)}(\theta)$ is $L^1(\pi)$ for $n \geq N$. Moreover $W_t(\theta) = \lim_{n\to\infty} W_t^{(n)}(\theta)$ given by (3.4) is also $L^1(\pi)$. Thus for sufficiently large n we use dominated convergence to show,

$$\begin{aligned}
\widehat{b}_{t,\lambda} &= \lim_{n\to\infty} \widehat{b}_{t,\lambda}^{(n)} \\
&= \lim_{n\to\infty} \frac{\int_{\Theta_\lambda} b(\theta, s_{\lfloor t \rfloor + 1}) W_t^{(n)}(\theta) d\pi(\theta)}{\int_{\Theta_\lambda} W_t^{(n)}(\theta) d\pi(\theta)} \\
&= \frac{\int_{\Theta_\lambda} b(\theta, s_{\lfloor t \rfloor + 1}) W_t(\theta) d\pi(\theta)}{\int_{\Theta_\lambda} W_t(\theta) d\pi(\theta)}.
\end{aligned}$$

Now let $\widehat{W}_{t,\lambda}$ be the wealth achieved by $\widehat{b}_{t,\lambda}$ up to time t. Again an application of the dominated convergence theorem shows,

$$\begin{aligned}
\widehat{W}_{t,\lambda} &= \lim_{n\to\infty} \widehat{W}_{t,\lambda}^{(n)} \\
&= \lim_{n\to\infty} \int_{\Theta_\lambda} W_t^{(n)}(\theta) d\pi(\theta) \\
&= \int_{\Theta_\lambda} W_t(\theta) d\pi(\theta).
\end{aligned}$$

The last step of the proof is to let $\lambda \to \infty$. Recall from equation (3.4) that $W_t(\theta)$ is a sum of exponential quadratic in $b(\theta, s)$. Moreover, because (by the minimal path assumptions) the matrices $\left\{K_\tau^\dagger\right\}_{\tau=1}^{\lfloor t\rfloor}$, $K_t^\dagger$ are positive definite, it is evident from (3.4) that this quadratic achieves a finite maximum on Θ. Thus for any $\lambda > 0$, $\max_{\theta\in\Theta_\lambda} W_t(\theta) \leq \max_{\theta\in\Theta} W_t(\theta) = W_t^* < \infty$. Furthermore, note that $W_t(\theta)1_{\theta\in\Theta_\lambda} \leq W_t^*$ for all λ and that $\lim_{\lambda\to\infty} W_t(\theta)1_{\theta\in\Theta_\lambda} = W_t(\theta)$ point-wise. By applying dominated convergence once more we show that,

$$
\begin{aligned}
\widehat{b}_t &= \lim_{\lambda\to\infty} \widehat{b}_{t,\lambda} \\
&= \lim_{\lambda\to\infty} \frac{\int_{\Theta_\lambda} b(\theta, s_{\lfloor t\rfloor+1}) W_t(\theta) d\pi(\theta)}{\int_{\Theta_\lambda} W_t(\theta) d\pi(\theta)} \\
&= \lim_{\lambda\to\infty} \frac{\int_{\Theta} b(\theta, s_{\lfloor t\rfloor+1}) W_t(\theta) \mathbf{1}_{\theta\in\Theta_\lambda} d\pi(\theta)}{\int_{\Theta} W_t(\theta) \mathbf{1}_{\theta\in\Theta_\lambda} d\pi(\theta)} \\
&= \frac{\int_{\Theta} b(\theta, s_{\lfloor t\rfloor+1}) W_t(\theta) d\pi(\theta)}{\int_{\Theta} W_t(\theta) d\pi(\theta)},
\end{aligned}
$$

and,

$$
\begin{aligned}
\widehat{W}_t &= \lim_{\lambda\to\infty} \widehat{W}_{t,\lambda} \\
&= \lim_{\lambda\to\infty} \int_{\Theta_\lambda} W_t(\theta) d\pi(\theta) \\
&= \lim_{\lambda\to\infty} \int_{\Theta} W_t(\theta) \mathbf{1}_{\theta\in\Theta_\lambda} d\pi(\theta) \\
&= \int_{\Theta} W_t(\theta) d\pi(\theta)..
\end{aligned}
$$

■

Looking at the form $\widehat{b}_t$, we see that at each instant t, portfolio $\widehat{b}_t$ invests in an average of the different strategies $b(\theta, s_\tau)$. Strategies are weighted in proportion to the wealth they have generated thus far. Thus, those that have done better in the past are weighted more heavily. Just as in the discrete case, $\widehat{b}_t$ can be equated with the practice of distributing initial wealth among a continuum of investment managers, each of whom uses a different θ to determine wealth allocations at each instance.

We now turn attention to proving that $\widehat{b}_t$ is universal with respect to the target wealth $W_t^* \equiv \max_{\theta\in\Theta} W_t(\theta)$. Actually we will prove something a bit stronger, namely that the ratio $W_t^*/\widehat{W}_t$ is bounded polynomially in t. We shall see that the order of these polynomial bounds corresponds nicely with those proven for discrete time.

The proof will require a few additional assumptions in addition to those of the minimal path and class conditions. The new assumptions regard bounding the size of the cumulative within-period empirical covariation and drift. In particular we will now assume that the sums of maximum eigenvalues (i.e. spectral radii) of the covariance matrices

$\sum_{\tau \in I(t)} \lambda_{\max}\left(K_t^{\dagger}\right)$ and sums of euclidean norms of drifts of $\sum_{\tau \in I(t)} \left\|\mu_\tau^{\dagger}\right\|$ are $O(t)$ (recall that $I(t)$ is the index set $\{1, \ldots, \lceil t \rceil - 1, t\}$). In addition to this, we will also need to assume that the parameter space Θ is now compact. These properties are stated rigorously in the following universality conditions:

General Universality Conditions

1. minimal path and class conditions hold.
2. $\Theta \subset \mathbf{R}^d$ is *compact* and convex.
3. there exists a constant $L_K > 0$ independent of t such that $\sum_{\tau \in I(t)} \lambda_{\max}\left(K_\tau^{\dagger}\right) < L_K \lceil t \rceil$, where $\lambda_{\max}\left(K_\tau^{\dagger}\right)$ denotes the maximum eigenvalue of $K_\tau^{\dagger}$.
4. there exists a constant $L_\mu > 0$ independent of t such that $\sum_{\tau \in I(t)} \|\mu_\tau^{\dagger}\| < L_\mu \lceil t \rceil$ for all $t > 0$.

To this point in our development we have always considered the parameter space of our target class and the parameter space over which we compute portfolio $\widehat{b}_t$ as one and the same. However, we will see shortly that advantages can be gained by viewing them as distinct spaces.

To clarify, suppose we consider a target class $(\Theta_A, S, b(\theta, s))$ satisfying the universality conditions. In an effort to be universal with respect to this class we might use the proposed universal procedure $\widehat{b}_t$ of equation (3.6). An implicit assumption used thus far is that when computing $\widehat{b}_t$, the domain of integration Θ in (3.6) is the same as the parameter space Θ_A of the target class. If this is the case, Theorem 3.3.2 given below proves that $\widehat{b}_t$ is universal. However, we also see from Theorem 3.3.2 that it isn't strictly necessarily for the domain of integration to be equal to Θ_A in order to get a universal procedure. If we can assume that the mapping $b : \Theta \times S \to B$ is extendable to a particular superset Θ_B of Θ_A then $\widehat{b}_t$ computed over all of Θ_B will also be universal with respect to target class $(\Theta_A, S, b(\theta, s))$.

The reason for wanting to consider larger domains of integration Θ_B is that the computation of $\widehat{b}_t$ can be radically simplified for specific choices of Θ_B, in particular for the choice $\Theta_B = \mathbf{R}^d$. Recall however that the universality conditions require that the parameter space Θ_A be a compact set of $\mathbf{R}^d$. Thus, if we were forced to use Θ_A as the domain of integration for $\widehat{b}_t$ we would never be in a position to exploit these nice computational properties. The freedom to set $\Theta_B = \mathbf{R}^d$ independently of Θ_A without losing the universality property will be of paramount importance in subsequent sections where we will prove the existence of easily computable universal procedures for certain types of linearly parameterized target classes.

Theorem 3.3.2 *Suppose that the target class $(\Theta_A, S, b(\theta, s))$ and price paths collectively satisfy the General Universality Conditions. Suppose also that the mapping $b : \Theta \times S \rightarrow B$ is extendible to some $\Theta_B \subseteq \mathbf{R}^d$, where Θ_B is any π-measurable superset of Θ_A. Let W_t^* be the target wealth associated with $(\Theta_A, S, b(\theta, s))$ and let $\widehat{W}_t$ be the wealth attained by $\widehat{b}_t$ applied to $(\Theta_B, S, b(\theta, s))$. Suppose π is absolutely continuous with respect to Lebesgue measure on Θ_B and has positive derivative on Θ_A. Then $\widehat{b}_t$ is universal with respect to $(\Theta_A, S, b(\theta, s))$ in the sense that,*

$$\frac{W_t^*}{\widehat{W}_t} = O(t^d).$$

Furthermore if,

1. *for some fixed $T > 0$ the interior of Θ_A holds an optimal parametrization θ_t^* for all $t \geq T$,*
2. *the components of $b(\theta, s)$ are uniformly bounded on $\Theta_A \times S$, and,*
3. *for each $s \in S$, $b(\theta, s)$ is differentiable on Θ_A with uniformly bounded first and second order partials,*

then,

$$\frac{W_t^*}{\widehat{W}_t} = O(t^{d/2}).$$

Proof. If $W_t^* = 0$ for some t the theorem follows immediately, so assume that $W_t^* > 0$ for all t. Define the set $\Theta_t^* = \{\theta \in \Theta_A : \|\theta - \theta_t^*\| \leq 1/\lceil t \rceil\}$ and let $\theta_t^\dagger \in \arg\min_{\theta \in \Theta_t^*} W_t(\theta)$. Then from Lemma 3.3.1,

$$\begin{aligned} \frac{\widehat{W}_t}{W_t^*} &= \int_{\Theta_B} \frac{W_t(\theta)}{W_t^*} d\pi(\theta) \\ &\geq \int_{\Theta_t^*} \frac{W_t(\theta)}{W_t^*} d\pi(\theta) \\ &\geq \frac{W_t(\theta_t^\dagger)}{W_t^*} \pi(\Theta_t^*). \end{aligned}$$

Substituting equation (3.4) for $W_t(\theta_t^\dagger)$ and W_t^* and using index set $I(t) = \{1, \ldots, \lceil t \rceil - 1, t\}$, we see that the above is equal to,

$$= \exp\left\{ \sum_{\tau \in I(t)} \left(\mu_\tau^\dagger\right)' \left(\widetilde{b}(\theta_t^\dagger, s_{\lceil\tau\rceil}) - \widetilde{b}(\theta_t^*, s_{\lceil\tau\rceil})\right) + \frac{1}{2} \sum_{\tau \in I(t)} \sum_{j=1}^{m} K_{\tau,j,j}^\dagger \left(b_j(\theta_t^\dagger, s_{\lceil\tau\rceil}) - b_j(\theta_t^*, s_{\lceil\tau\rceil})\right) \right.$$
$$\left. -\frac{1}{2} \sum_{\tau \in I(t)} \left(\widetilde{b}(\theta_t^\dagger, s_{\lceil\tau\rceil}) - \widetilde{b}(\theta_t^*, s_{\lceil\tau\rceil})\right)' K_\tau^\dagger \left(\widetilde{b}(\theta_t^\dagger, s_{\lceil\tau\rceil}) - \widetilde{b}(\theta_t^*, s_{\lceil\tau\rceil})\right) \right\} \pi(\Theta_t^*). \tag{3.15}$$

We bound each term using the universality assumptions. Note that,

$$\begin{aligned}
\sum_{\tau \in I(t)} \left| \left(\mu_\tau^\dagger\right)' \left(\widetilde{b}(\theta_t^\dagger, s_\tau) - \widetilde{b}(\theta_t^*, s_\tau)\right) \right| &\le \sum_{\tau \in I(t)} \left\|\mu_\tau^\dagger\right\| \left\|\widetilde{b}(\theta_t^\dagger, s_{\lceil\tau\rceil}) - \widetilde{b}(\theta_t^*, s_{\lceil\tau\rceil})\right\| \\
&\le \sum_{\tau \in I(t)} \left\|\mu_\tau^\dagger\right\| L_b \|\theta_t^\dagger - \theta_t^*\| \\
&\le \sum_{\tau \in I(t)} \left\|\mu_{t'}^\dagger\right\| L_b / \lceil t \rceil \quad \text{(by virtue of the definition of } \Theta_t^*) \\
&\le L_\mu \lceil t \rceil L_b / \lceil t \rceil \quad \text{by universality conditions.} \\
&\le L_\mu L_b. \qquad (3.16)
\end{aligned}$$

Similarly, use the fact that any element of a positive definite matrix is bounded by its spectral radius to show that,

$$\begin{aligned}
\left| \sum_{\tau \in I(t)} \sum_{j=1}^{m} K_{\tau,j,j}^\dagger \left(b_j(\theta_t^\dagger, s_{\lceil\tau\rceil}) - b_j(\theta_t^*, s_{\lceil\tau\rceil})\right) \right| &\le \sum_{\tau \in I(t)} \left\| \operatorname{diag}\left(K_\tau^\dagger\right) \right\| \left\| b(\theta_t^\dagger, s_{\lceil\tau\rceil}) - b(\theta_t^*, s_{\lceil\tau\rceil}) \right\| \\
&\le \sum_{\tau \in I(t)} \sqrt{m} \lambda_{\max} \left(K_\tau^\dagger\right) L_b \|\theta_t^\dagger - \theta_t^*\| \\
&\le \sqrt{m} L_b / \lceil t \rceil \sum_{\tau \in I(t)} \lambda_{\max} \left(K_\tau^\dagger\right) \\
&\le \sqrt{m} \left(L_b / \lceil t \rceil\right) L_K \lceil t \rceil \\
&= \sqrt{m} L_b L_K. \qquad (3.17)
\end{aligned}$$

Finally, for the quadratic term,

$$
\begin{aligned}
&\sum_{\tau \in I(t)} \left(\widetilde{b}(\theta_t^\dagger, s_{\lceil\tau\rceil}) - \widetilde{b}(\theta_t^*, s_{\lceil\tau\rceil})\right)' K_\tau^\dagger \left(\widetilde{b}(\theta_t^\dagger, s_{\lceil\tau\rceil}) - \widetilde{b}(\theta_t^*, s_{\lceil\tau\rceil})\right) \\
&\le \sum_{\tau \in I(t)} \lambda_{\max}\left(K_\tau^\dagger\right) \left\|\widetilde{b}(\theta_t^\dagger, s_{\lceil\tau\rceil}) - \widetilde{b}(\theta_t^*, s_{\lceil\tau\rceil})\right\|^2 \\
&\le \sum_{\tau \in I(t)} \lambda_{\max}\left(K_\tau^\dagger\right) L_b^2 \|\theta_t^\dagger - \theta_t^*\|^2 \\
&\le L_b^2 / \lceil t \rceil^2 \sum_{\tau \in I(t)} \lambda_{\max}\left(K_\tau^\dagger\right) \\
&\le L_b^2 / \lceil t \rceil^2 L_K \lceil t \rceil \\
&= L_b^2 L_K / \lceil t \rceil \,. \qquad (3.18)
\end{aligned}
$$

Substituting (3.16), (3.17), and (3.18) into (3.15) we further bound the wealth ratio by,

$$
\begin{aligned}
\frac{\widehat{W}_t}{W_t^*} &\ge \exp\left\{-L_\mu L_b - \frac{\sqrt{m}}{2} L_K L_b - \frac{1}{2} L_K L_b^2 / \lceil t \rceil\right\} \pi(\Theta_t^*) \\
&\ge \exp\left\{-L_b\left(L_\mu + \frac{\sqrt{m}}{2} L_K + \frac{1}{2} L_K L_b\right)\right\} \pi(\Theta_t^*).
\end{aligned}
$$

Lemma 5.2.1 of the appendix shows that for some constant $R > 0$, we can bound $\pi(\Theta_t^*) \ge R \lceil t \rceil^{-d}$ for any $t > 0$. Thus,

$$
\frac{\widehat{W}_t}{W_t^*} \ge \exp\left\{-L_b\left(L_\mu + \frac{\sqrt{m}}{2} L_K + \frac{1}{2} L_K L_b\right)\right\} R \lceil t \rceil^{-d}.
$$

Upon inverting this ratio we conclude that $\frac{W_t^*}{\widehat{W}_t}$ is $O(t^d)$.

To improve this bound to $O(t^{d/2})$ we assume that there exists a T and maximizers $\{\theta_t^*\}_{t \ge T}$ such that $\theta_t^* \in int(\Theta_A)$ for all $t \ge T$. Also we assume that for all $s \in S$ and $\theta = (\phi_1, \ldots, \phi_d) \in \Theta_A$ there exists $M_0 > 0$, $M_1 > 0$ and $M_2 > 0$ such that,

$$
|b_j(\theta, s)| < M_0,
$$

$$
\left|\frac{\partial b_j(\theta, s)}{\partial \phi_r}\right| < M_1,
$$

and,

$$
\left|\frac{\partial^2 b_j(\theta, s)}{\partial \phi_r \partial \phi_s}\right| < M_2
$$

for any $1 \le j \le m$, and $1 \le r, s \le d$.

Since $W_t(\theta)$ is infinitely differentiable with respect to the twice differentiable function $b(\theta, s)$, $W_t(\theta)$ is twice differentiable with respect to θ. Similarly, $l_t(\theta) \equiv \log W_t(\theta)$ is also twice differentiable. Thus,

$$\begin{aligned} l_t(\theta) &= l_t(\theta_t^*) + \sum_{r=1}^{d}(\theta_r - \theta_{t,r}^*)\frac{\partial l_t(\theta_t^*)}{\partial \phi_r} \\ &+\frac{1}{2}\sum_{r,s=1}^{d}(\theta_r - \theta_{t,r}^*)(\theta_s - \theta_{t,s}^*)\frac{\partial^2 l_t(\theta_t')}{\partial\phi_r\partial\phi_s}, \end{aligned}$$

for some $\theta_t' \in \Theta$ between θ and $\theta_{t,}^*$. For $t \geq T$ recall that θ_t^* maximizes $W_t(\theta)$ in the interior of Θ_A. This necessitates that,

$$\frac{\partial W_t(\theta_t^*)}{\partial \phi_r} = 0,$$

which in turn implies,

$$\frac{\partial l_t(\theta_t^*)}{\partial \phi_r} = \frac{1}{W_t(\theta^*)}\frac{\partial W_t(\theta_t^*)}{\partial \phi_r} = 0.$$

In light of the above we write,

$$l_t(\theta) = l_t(\theta_t^*) + \frac{1}{2}\sum_{r,s=1}^{d}(\theta_r - \theta_{t,r}^*)(\theta_s - \theta_{t,s}^*)\frac{\partial^2 l_t(\theta_t')}{\partial\phi_r\partial\phi_s}.$$

Now we aim to bound the second order partial of $l_t(\theta)$. Note that,

$$\begin{aligned} \left|\frac{\partial^2}{\partial\phi_r\partial\phi_s}l_t(\theta)\right| &= \left|\sum_{\tau\in I(t)}\frac{\partial^2}{\partial\phi_r\partial\phi_s}\left(\mu_\tau^\dagger\right)'\tilde{b}(\theta, s_{\lceil\tau\rceil}) + \frac{1}{2}\sum_{\tau\in I(t)}\frac{\partial^2}{\partial\phi_r\partial\phi_s}\sum_{j=1}^{m}K_{\tau,j,j}^\dagger b_j(\theta, s_{\lceil\tau\rceil})\right. \\ &\left.-\frac{1}{2}\sum_{\tau\in I(t)}\frac{\partial^2}{\partial\phi_r\partial\phi_s}\left[\tilde{b}(\theta, s_{\lceil\tau\rceil})'K_\tau^\dagger\tilde{b}(\theta, s_{\lceil\tau\rceil})\right]\right|. \end{aligned} \tag{3.19}$$

We aim to bound each of the terms in the above expression. Our first task is to use the assumptions that $\sum_{\tau\in I(t)}\left\|\mu_\tau^\dagger\right\| < L_\mu\lceil t\rceil$ and $\left|\frac{\partial^2 b_j(\theta,s)}{\partial\phi_r\partial\phi_s}\right| < M_2$ to establish the bound,

$$\begin{aligned} \sum_{\tau\in I(t)}\left|\frac{\partial^2}{\partial\phi_r\partial\phi_s}\left(\mu_\tau^\dagger\right)'\tilde{b}(\theta, s_{\lceil\tau\rceil})\right| &= \sum_{\tau\in I(t)}\left|\mu_\tau^\dagger\cdot\frac{\partial^2}{\partial\phi_r\partial\phi_s}\tilde{b}(\theta, s_{\lceil\tau\rceil})\right| \\ &\leq \sum_{\tau\in I(t)}\left\|\mu_\tau^\dagger\right\|\sum_{j=1}^{m}\left|\frac{\partial^2}{\partial\phi_r\partial\phi_s}b_j(\theta, s_{\lceil\tau\rceil})\right| \\ &\leq \sum_{\tau\in I(t)}\left\|\mu_\tau^\dagger\right\|\sum_{j=1}^{m}M_2 \\ &\leq mM_2\sum_{\tau\in I(t)}\left\|\mu_\tau^\dagger\right\| \\ &\leq mM_2L_\mu\lceil t\rceil. \end{aligned} \tag{3.20}$$

Recall again that the positive definiteness of $K_\tau^\dagger$ insures that each of the diagonal elements $K_{\tau,j,j}^\dagger$ is bounded above by spectral radius $\lambda_{\max}(K_\tau^\dagger)$. Hence we note that,

$$\begin{aligned}\sum_{\tau\in I(t)}\left|\frac{\partial^2}{\partial\phi_r\partial\phi_s}\frac{1}{2}\sum_{j=1}^{m}K_{\tau,j,j}^\dagger b_j(\theta,s_{\lceil\tau\rceil})\right| &\leq \frac{1}{2}\sum_{\tau\in I(t)}\sum_{j=1}^{m}K_{\tau,j,j}^\dagger\left|\frac{\partial^2}{\partial\phi_r\partial\phi_s}b_j(\theta,s_{\lceil\tau\rceil})\right|\\ &\leq \frac{1}{2}\sum_{\tau\in I(t)}\sum_{j=1}^{m}\lambda_{\max}(K_\tau^\dagger)M_2\\ &\leq \frac{1}{2}mM_2\sum_{\tau\in I(t)}\lambda_{\max}(K_\tau^\dagger)\\ &\leq \frac{1}{2}mM_2L_K\lceil t\rceil.\end{aligned}\tag{3.21}$$

Finally for the quadratic term,

$$\begin{aligned}&\sum_{\tau\in I(t)}\left|-\frac{1}{2}\frac{\partial^2}{\partial\phi_r\partial\phi_s}\left[\widetilde{b}(\theta,s_{\lceil\tau\rceil})'K_\tau^\dagger\widetilde{b}(\theta,s_{\lceil\tau\rceil})\right]\right|\\ \leq\;&\sum_{\tau\in I(t)}\left|\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}K_{\tau,i,j}^\dagger\frac{\partial^2}{\partial\theta_r\partial\theta_s}\left(b_i(\theta,s_{\lceil\tau\rceil})b_j(\theta,s_{\lceil\tau\rceil})\right)\right|\\ \leq\;&\frac{1}{2}\sum_{\tau\in I(t)}\sum_{i=1}^{m}\sum_{j=1}^{m}\left|K_{\tau,i,j}^\dagger\right|\left|2\left(\frac{\partial b_i(\theta,s_{\lceil t'\rceil})}{\partial\theta_r}\frac{\partial b_j(\theta,s_{\lceil t'\rceil})}{\partial\theta_s}\right)+2\left(b_i(\theta,s_{\lceil t'\rceil})\frac{\partial^2 b_j(\theta,s_{\lceil t'\rceil})}{\partial\theta_r\partial\theta_s}\right)\right|.\end{aligned}$$

Now use the property of positive definite matrices that $\left|K_{\tau,i,j}^\dagger\right|\leq\lambda_{\max}\left(K_\tau^\dagger\right)$ for any entry of the matrix.

$$\begin{aligned}\leq\;&\frac{1}{2}\sum_{\tau\in I(t)}\lambda_{\max}\left(K_\tau^\dagger\right)\sum_{i=1}^{m}\sum_{j=1}^{m}\left|2\left(\frac{\partial b_i(\theta,s_{\lceil t'\rceil})}{\partial\theta_r}\frac{\partial b_j(\theta,s_{\lceil t'\rceil})}{\partial\theta_s}\right)+2\left(b_i(\theta,s_{\lceil t'\rceil})\frac{\partial^2 b_j(\theta,s_{\lceil t'\rceil})}{\partial\theta_r\partial\theta_s}\right)\right|\\ \leq\;&\frac{1}{2}\sum_{\tau\in I(t)}\lambda_{\max}\left(K_\tau^\dagger\right)\sum_{i=1}^{m}\sum_{j=1}^{m}\left(2M_1^2+2M_0M_2\right)\\ \leq\;&L_K\lceil t\rceil m^2\left(M_1^2+M_0M_2\right).\end{aligned}$$

Thus we conclude that

$$\sum_{\tau\in I(t)}\left|-\frac{1}{2}\frac{\partial^2}{\partial\phi_r\partial\phi_s}\left[\widetilde{b}(\theta,s_{\lceil t'\rceil})'K_{t'}^\dagger\widetilde{b}(\theta,s_{\lceil t'\rceil})\right]\right|\leq m^2\left(M_1^2+M_0M_2\right)L_K\lceil t\rceil.\tag{3.22}$$

Now set,

$$M\equiv mM_2L_\mu+\frac{1}{2}mM_2L_K+m^2\left(M_1^2+M_0M_2\right)L_K.$$

Using bounds (3.20),(3.21), and (3.22) on (3.19) we conclude that,

$$\left|\frac{\partial^2}{\partial\phi_r\partial\phi_s}l_t(\theta)\right| \le M\lceil t\rceil,$$

Now define the set $\Theta^*_{\sqrt{t}} \equiv \{\theta \in \Theta_A | \|\theta - \theta^*_t\| \le 1/\sqrt{\lceil t\rceil}\}$. Returning to log wealth $l_t(\theta)$, we note that for $\theta \in \Theta^*_{\sqrt{t}}$,

$$\begin{aligned} l_t(\theta) &= l_t(\theta^*_t) + \frac{1}{2}\sum_{r,s=1}^{d}(\theta_r - \theta^*_{t,r})(\theta_s - \theta^*_{t,s})\frac{\partial^2 l_t(\theta'_t)}{\partial\theta_r\partial\theta_r} \\ &\ge l_t(\theta^*_t) - \frac{1}{2}\sum_{r,s=1}^{d}\|\theta_r - \theta^*_{t,r}\|\,\|\theta_s - \theta^*_{t,s}\|\,M\lceil t\rceil, \end{aligned}$$

and since $\|\theta_r - \theta^*_{t,r}\| \le 1/\sqrt{\lceil t\rceil}$ on $\Theta^*_{\sqrt{t}}$ it follows that,

$$l_t(\theta) \ge l_t(\theta^*_t) - \frac{1}{2}d^2M.$$

Thus in the neighborhood of $\Theta^*_{\sqrt{t}}$,

$$W_t(\theta) \ge \exp\{l_t(\theta^*_t) - \frac{1}{2}d^2M\} = W^*_t\exp\{-\frac{1}{2}d^2M\}.$$

Redefine, $\theta^\dagger_t \in \arg\min_{\theta\in\Theta^*_{\sqrt{t}}} W_t(\theta)$ and note that

$$\begin{aligned} \frac{\widehat{W}_t}{W^*_t} &= \int_{\Theta_B}\frac{W_t(\theta)}{W^*_t}d\pi(\theta) \\ &\ge \int_{\Theta^*_{\sqrt{t}}}\frac{W_t(\theta)}{W^*_t}d\pi(\theta) \\ &\ge \frac{W_t(\theta^\dagger_t)}{W^*_t}\pi(\Theta^*_{\sqrt{t}}) \\ &\ge \exp\{-\frac{1}{2}d^2M\}\pi(\Theta^*_{\sqrt{t}}). \end{aligned}$$

Again we show in Lemma 5.2.1 of the appendix that there exists a constant $R > 0$ such that $\pi(\Theta^*_{\sqrt{t}}) \ge R\lceil\sqrt{t}\rceil^{-d}$. Thus

$$\frac{\widehat{W}_t}{W^*_t} \ge \exp\{-\frac{1}{2}d^2M\}R\left\lceil\sqrt{t}\right\rceil^{-d}.$$

Upon inverting the ratio we conclude that $\frac{W^*_t}{\widehat{W}_t}$ is $O(t^{d/2})$. ■

3.4 Universal Portfolios for Linear Classes

3.4.1 Introduction to Linear Classes

Recall that one of our main reasons for looking at universal procedures in continuous time is that we hoped to find instances for which the computation of the universal portfolio would be simplified. For this reason we now consider classes which are linearly parameterized in θ. Specifically we consider classes which use side information s_τ at the start of period τ to determine an $m \times d$ matrix $A_\tau = \dot{A}(s_t)$ used to set the stock portfolio,

$$\widetilde{b}(\theta, s_\tau) = (b_1(\theta, s_\tau), \ldots, b_m(\theta, s_\tau)) = A_\tau \theta.$$

As always the cash component $b_0(\theta, s_\tau)$ for such a portfolio would be set to $1 - \sum_{j=1}^{m} b_j(\theta, s_\tau)$.

These linear classes include many of the class types we have considered so far. For example, the class discussed in section 3.2.2 is an example of a linear class. Also, the family of constant rebalanced portfolios would be another example of a linear class.

The wealth of these classes, $W_t(\theta)$, is entirely expressible in terms of quantities depending only on the side information matrices $A(s)$, the empirical covariance matrices $K_\tau^\dagger$, and empirical drifts $\mu_\tau^\dagger$. Also, we find that $W_t(\theta)$ now yields a unique closed form solution for the parameter achieving maximum wealth, i.e. $\theta_t^* = \arg\max_{\theta\in\Theta} W_t(\theta)$. These two points are proven in the following lemma.

Lemma 3.4.1 *Suppose that* $\widetilde{b}(\theta, s_\tau) = A_\tau\theta$ *and that the* $d \times d$ *matrix,*

$$\Omega_t^{-1} \equiv A_1' K_1^\dagger A_1 + \cdots + A_{\lceil t\rceil-1}' K_{\lceil t\rceil-1}^\dagger A_{\lceil t\rceil-1} + A_{\lceil t\rceil}' K_t^\dagger A_{\lceil t\rceil}, \tag{3.23}$$

is invertible. Define,

$$u_t \equiv A_{\lceil t\rceil}' \left[\mu_t^\dagger + \frac{1}{2}\mathrm{diag} K_t^\dagger\right], \tag{3.24}$$

and,

$$v_t \equiv \Omega_t(u_1 + \cdots + u_{\lceil t\rceil-1} + u_t). \tag{3.25}$$

The wealth achieved by $b(\theta, s_\tau) = \left(1 - \sum_{j=1}^{m} b_j(\theta, s_\tau), \widetilde{b}(\theta, s_\tau)\right)$ *by the end of period* T *is,*

$$W_t(\theta) = W_0 \exp\left\{-\frac{1}{2}(\theta - v_t)'\Omega_t^{-1}(\theta - v_t) + \frac{1}{2}v_t'\Omega_t^{-1}v_t\right\} \tag{3.26}$$

Furthermore, the unique optimal parameter $\theta_t^* = \arg\max_{\theta\in\Theta} W_t(\theta)$ *achieving* W_t^* *(the best wealth in hindsight) is given by,*

$$\theta_t^* = v_t,$$

and thus,

$$W_t^* = W_0 \exp\left\{\frac{1}{2}v_t'\Omega_t^{-1}v_t\right\}. \tag{3.27}$$

Proof. Starting with the exponent of (3.26) note that,

$$\begin{aligned}
& -\frac{1}{2}(\theta - v_t)'\Omega_t^{-1}(\theta - v_t) + \frac{1}{2}v_t\Omega_t^{-1}v_t \\
= & -\frac{1}{2}\theta'\Omega_t^{-1}\theta + \theta'\Omega_t^{-1}v_T - \frac{1}{2}v_t'\Omega_t^{-1}v_t + \frac{1}{2}v_t'\Omega_t^{-1}v_t \\
= & -\frac{1}{2}\theta'\Omega_t^{-1}\theta' + \theta'\Omega_t^{-1}v_t \\
= & -\frac{1}{2}\sum_{\tau\in I(t)} \theta' A_\tau' K_\tau^\ddagger A_\tau\theta + \theta'\Omega_t^{-1}\Omega_t \sum_{\tau\in I(t)} u_\tau \\
= & -\frac{1}{2}\sum_{\tau\in I(t)} \theta' A_\tau' K_\tau^\ddagger A_\tau\theta + \theta' \sum_{\tau\in I(t)} u_\tau \\
= & -\frac{1}{2}\sum_{\tau\in I(t)} \theta' A_\tau' K_\tau^\ddagger A_\tau\theta + \theta' \sum_{\tau\in I(t)} A_\tau' \left[\mu_\tau^\dagger + \frac{1}{2}\text{diag}K_\tau^\ddagger\right] \\
= & -\frac{1}{2}\sum_{\tau\in I(t)} \widetilde{b}'(\theta, s_\tau) K_\tau^\ddagger \widetilde{b}(\theta, s_\tau) + \sum_{\tau\in I(t)} \widetilde{b}'(\theta, s_\tau)\left[\mu_\tau + \frac{1}{2}\text{diag}K_\tau^\ddagger\right].
\end{aligned}$$

This is the same as the exponent of (3.4) so it follows that (3.26) holds.

To prove $\theta_t^* = v_t$ note that wealth is maximized when $-\frac{1}{2}(\theta - v_t)'\Omega_t^{-1}(\theta - v_t)$ is maximized. Since Ω^{-1} is the sum of positive definite matrices, it is itself a positive definite matrix. Thus the quadratic is non-positive for all θ .and is maximized when $(\theta - v_t) = 0$. Thus the unique maximizer is $\theta_t^* = v_t$. ■

Now we seek a universal portfolio for these linear classes. Since these classes are subclasses of the portfolio families considered in the previous section we immediately know that portfolio (3.6) of Theorem 3.3.2 is a universal portfolio with respect to the linear class $(\Theta, S, A(s)\theta)$. Let us consider in further detail what happens when we adapt portfolio (3.6) to the present case. Clearly it is telling us to trade at each instance t the portfolio $\widehat{b}_t = \left(\widehat{b}_{t,0}, \widehat{\widehat{b}}_t\right)$, where

$$\widehat{\widehat{b}}_t = \frac{\int_\Theta A_{\lfloor t\rfloor+1}\theta W_t(\theta)d\pi(\theta)}{\int_\Theta W_t(\theta)d\pi(\theta)} \tag{3.28}$$

represents the proportions put in stocks and $\widehat{b}_{t,0} = 1 - \sum_{j=1}^m \widehat{b}_{t,j}$ represents the proportion put in cash (for clarity, $\widehat{\widehat{b}}_t = \left(\widehat{b}_{t,1}, \ldots, \widehat{b}_{t,m}\right)$). By Lemma 3.3.1 $\widehat{b}_t$ achieves wealth,

$$\widehat{W}_t = \int_\Theta W_t(\theta)d\pi(\theta). \tag{3.29}$$

In order to have easy computation of $\widehat{b}_t$ it would help if we could find a choice of Θ and $\pi(\theta)$ that yielded nice closed form expressions for (3.28). Here the Bayesian concept of a normal conjugate prior will prove to be very useful. This concept refers to the fact that if you apply a Gaussian prior to a Gaussian sampling density, the resulting posterior is also Gaussian. Lets adapt this concept to the present case. Recall from (3.26) that $W_t(\theta)$ is an exponential quadratic in θ. Thus $W_t(\theta)$ is equivalent to some non-normalized Gaussian density. Suppose now that we choose Θ to be $\mathbf{R}^d$ and that we choose π to be some arbitrary Gaussian measure on $\mathbf{R}^d$. Then from the property of normal self-conjugation, the measure

$$\frac{W_t(\theta)d\pi(\theta)}{\int_\Theta W_t(\theta)d\pi(\theta)}$$

is also Gaussian. Thus it follows from the form of (3.28) that the calculation of $\widehat{b}_t$ is essentially equivalent to a normal expectation calculation! Such calculations are easily computable and yield simple solutions. The following lemma gives a precise formula for the stock portion of $\widehat{b}_t$ and its corresponding wealth.

Lemma 3.4.2 *Upon setting $\Theta = \mathbf{R}^d$ and $\pi \sim N(\lambda, \Lambda)$, the stock portion of universal procedure $\widehat{b}_t$ given by*

$$\widetilde{\widetilde{b}}_t = \frac{\int_\Theta A_{\lfloor t\rfloor+1}\theta W_t(\theta)d\pi(\theta)}{\int_\Theta W_t(\theta)d\pi(\theta)}$$

can be rewritten as,

$$\widetilde{\widetilde{b}}_t = A_{\lfloor t\rfloor+1}\Psi_t \text{ where } \Psi_t = \left(\Omega_t^{-1} + \Lambda^{-1}\right)^{-1}\left(\sum_{\tau\in I(t)} u_\tau + \Lambda^{-1}\lambda\right) \tag{3.30}$$

for Ω_t^{-1} and u_τ as given in (3.23) and (3.24) respectively. Similarly, the wealth achieved by $\widehat{b}_t$ is writable as,

$$\widehat{W}_\tau = W_0\left|\Lambda\right|^{-1/2}\left|\Omega_t^{-1} + \Lambda^{-1}\right|^{-1/2}\exp\left\{\frac{1}{2}\left[\Psi_t'\left(\Omega_t^{-1}+\Lambda^{-1}\right)\Psi_t - \lambda'\Lambda^{-1}\lambda\right]\right\}. \tag{3.31}$$

Proof. We prove the latter statement first. From Lemma 3.3.1,

$$\widehat{W}_\tau = \int_{\mathbf{R}^d} W_t(\theta)d\pi(\theta).$$

Setting $W_t(\theta)$ as in (3.4) and using $\pi \sim N(\Lambda, \lambda)$, we see that $\widehat{W}_\tau$ equals,

$$\begin{aligned}
\widehat{W}_\tau &= \int_{\mathbf{R}^d} W_0 \exp\left\{-\frac{1}{2}\sum_{\tau\in I(t)} \theta' A'_\tau K^\dagger_\tau A_\tau \theta + \sum_{\tau\in I(t)} \theta' A'_\tau \left[\mu_\tau + \frac{1}{2}\mathrm{diag}K^\dagger_\tau\right]\right\} \times \\
&\quad |2\pi\Lambda|^{-1/2} \exp\left\{-\frac{1}{2}(\theta-\lambda)'\Lambda^{-1}(\theta-\lambda)\right\} d\theta. \\
&= W_0\, |2\pi\Lambda|^{-1/2} \int_{\mathbf{R}^d} \exp\left\{-\frac{1}{2}\theta'\Omega_t^{-1}\theta + \theta' \sum_{\tau\in I(t)} u_\tau \right. \\
&\quad \left. -\frac{1}{2}(\theta-\lambda)'\Lambda^{-1}(\theta-\lambda)\right\} d\theta \\
&= W_0\, |2\pi\Lambda|^{-1/2} \int_{\mathbf{R}^d} \exp\left\{-\frac{1}{2}\theta' \left(\Omega_t^{-1} +' \Lambda^{-1}\right)\theta \right. \\
&\quad \left. +\theta' \left(\sum_{\tau\in I(t)} u_\tau + \Lambda^{-1}\lambda\right) - \frac{1}{2}\lambda'\Lambda^{-1}\lambda\right\} d\theta
\end{aligned}$$

After rearranging terms and setting $\Psi_t = \left(\Omega_t^{-1} + \Lambda^{-1}\right)^{-1} \left(\sum_{\tau\in I(t)} u_\tau + \Lambda^{-1}\lambda\right)$ the above becomes,

$$\begin{aligned}
\widehat{W}_\tau &= W_0\, |2\pi\Lambda|^{-1/2} \int_{\mathbf{R}^d} \exp\left\{-\frac{1}{2}(\theta-\Psi_t)' \left(\Omega_t^{-1} + \Lambda^{-1}\right)(\theta-\Psi_t) \right. \\
&\quad \left. +\frac{1}{2}\Psi'_t \left(\Omega_t^{-1} + \Lambda^{-1}\right)\Psi_t - \frac{1}{2}\lambda'\Lambda^{-1}\lambda\right\} d\theta.
\end{aligned} \tag{3.32}$$

The integrand is in the form of a non-normalized Gaussian density which integrates to,

$$\int_{\mathbf{R}^d} \exp\left\{-\frac{1}{2}(\theta-\Psi_t)' \left(\Omega_t^{-1} + \Lambda^{-1}\right)(\theta-\Psi_t)\right\} d\theta = \left|2\pi \left(\Omega_t^{-1} + \Lambda^{-1}\right)^{-1}\right|^{1/2}.$$

Thus, we conclude that

$$\widehat{W}_\tau = W_0 \left(|\Lambda|\, |\Omega_t^{-1} + \Lambda^{-1}|\right)^{-1/2} \exp\left\{\frac{1}{2}\Psi'_t \left(\Omega_t^{-1} + \Lambda^{-1}\right)\Psi_t - \frac{1}{2}\lambda'\Lambda^{-1}\lambda\right\}.$$

To show $\widetilde{\widetilde{b}}_t = A_{\lfloor t\rfloor+1}\Psi_t$ we start with (3.28)and (3.32) and conclude with the simple argument that,

$$\begin{aligned}
\widetilde{\widetilde{b}}_t &= \frac{\int_{\mathbf{R}^d} A_{\lfloor t\rfloor+1}\theta \exp\left\{-\frac{1}{2}(\theta-\Psi_t)' \left(\Omega_t^{-1} + \Lambda^{-1}\right)(\theta-\Psi_t)\right\} d\theta}{\int_{\mathbf{R}^d} \exp\left\{-\frac{1}{2}(\theta-\Psi_t)' \left(\Omega_t^{-1} + \Lambda^{-1}\right)(\theta-\Psi_t)\right\} d\theta} \\
&= \frac{A_{\lfloor t\rfloor+1} \int_{\mathbf{R}^d} \theta \exp\left\{-\frac{1}{2}(\theta-\Psi_t)' \left(\Omega_t^{-1} + \Lambda^{-1}\right)(\theta-\Psi_t)\right\} d\theta}{\int_{\mathbf{R}^d} \exp\left\{-\frac{1}{2}(\theta-\Psi_t)' \left(\Omega_t^{-1} + \Lambda^{-1}\right)(\theta-\Psi_t)\right\} d\theta} \\
&= A_{\lfloor t\rfloor+1}\Psi_t.
\end{aligned}$$

■

Clearly the portfolio strategy $\widehat{b}_t$ of Lemma 3.4.2 has nice computational properties. If the current values of Ω_t^{-1} and $\sum_{\tau\in I(t)} u_\tau$ are available, the portfolio allocations and accumulated wealth at that instance are easily computed via simple linear calculations. As we will be repeatedly referring to this portfolio procedure it will be useful to have it explicitly defined. Hence we define,

Procedure 1 *Invest according to* $\widehat{b}_t = \left(\widehat{b}_{t,0}, \widehat{\widetilde{b}}_t\right)$ *, where stock component* $\widehat{\widetilde{b}}_t$ *is computed directly from,*

$$\widehat{\widetilde{b}}_t = \frac{\int_\Theta A_{\lfloor t\rfloor+1}\theta W_t(\theta)d\pi(\theta)}{\int_\Theta W_t(\theta)d\pi(\theta)} = A_{\lfloor t\rfloor+1}\Psi_t, \quad \text{where } \pi \sim N(\lambda, \Lambda).$$

with,

$$\begin{aligned} \Psi_t &= \left(\Omega_t^{-1} + \Lambda^{-1}\right)^{-1}\left(\sum_{\tau\in I(t)} u_\tau + \Lambda^{-1}\lambda\right), \\ \Omega_t^{-1} &\equiv A_1' K_1^\dagger A_1 + \cdots + A_{\lceil t\rceil-1}' K_{\lceil t\rceil-1}^\dagger A_{\lceil t\rceil-1} + A_{\lceil t\rceil}' K_t^\dagger A_{\lceil t\rceil}, \\ u_t &\equiv A_{\lceil t\rceil}'\left[\mu_t^\dagger + \frac{1}{2}\text{diag}K_t^\dagger\right]. \end{aligned}$$

The wealth achieved by $\widehat{b}_t$ *is given by equation (3.31) of Lemma 3.4.2.*

3.4.2 A Universality Theorem for Linear Classes

It is clear from Theorem 3.3.2 that Procedure 1 is universal with respect to the linear target class $(\Theta, S, A(s)\theta)$ *if* the General Universality Conditions hold. Recall that one of these conditions is that the parameter space Θ be compact. What we would like to do in this section is find conditions under which universality can be maintained when we increase the size of the parameter space to all of $\mathbf{R}^d$.

Although the General Universality Conditions aren't sufficiently strong to give universality for parameter space $\Theta = \mathbf{R}^d$, they along with Theorem 3.3.2 show that a universal procedure satisfying the theorem must come within a $O\left(n^d\right)$ factor of the wealth of any strategy $\theta \in \mathbf{R}^d$. The problem here is that the constant multiplying this order bound is dependent on the distance between θ and the origin. With regards to a universal procedure tracking optimal wealth $W(\theta_t^*)$, if the sequence of maximizing parameters θ_t^* tends to infinity with t then we will be unable to get a constant for the order bound that holds uniformly over t. This is the primary obstacle that must be overcome to get universality on $\Theta = \mathbf{R}^d$.

One way to escape this problem is to find conditions under which the sequence of θ_t^* is guaranteed to stay within a compact neighborhood. Once this is done, universality over all of $\mathbf{R}^d$ becomes a realistic goal. With this in mind we now present the following universality conditions for a linear class.

Alternate Universality Conditions (for a Linear Class) Assume the following conditions hold:

1. The minimal path conditions hold.
2. There exists a constant $L_\mu > 0$ independent of t such that $\sum_{\tau \in I(t)} \|\mu_\tau^\dagger\| < L_\mu \lceil t \rceil$ for all $t > 0$. (Recall that $I(t) = \{1, \ldots, \lceil t \rceil - 1, t\}$).
3. There exists a constant $L_K > 0$ independent of t such that $\sum_{\tau \in I(t)} \lambda_{\max}\left(K_\tau^\dagger\right) < L_K \lceil t \rceil$, where $\lambda_{\max}\left(K_\tau^\dagger\right)$ denotes the maximum eigenvalue of $K_\tau^\dagger$.
4. For any $s \in S$, the $m \times d$ matrix $A(s)$ has full rank and there exists a constant $L_A > 0$ independent of period s such that $\lambda_{\max}(A(s)\,A'(s)) \leq L_A$.
5. There exists some integer β (possibly depending on the price path and side information sequence) such that Ω_t^{-1} is invertible for all $t > \beta$.
6. For $t > \beta + 1$, there exists positive constants L_Ω^- and L_Ω^+ independent of t such the minimum and maximum eigenvalues of Ω_t^{-1}satisfy, $L_\Omega^- (\lceil t \rceil - 1) \leq \lambda_{\min}\left(\Omega_t^{-1}\right) \leq \lambda_{\max}\left(\Omega_t^{-1}\right) \leq L_\Omega^+ \lceil t \rceil$.

We should briefly comment on the fifth universality condition regarding the invertibilty of Ω_t^{-1}. If the dimensionality d of our parameter space is greater than the number of stocks m we can see from the form of Ω_t^{-1}, i.e.,

$$\Omega_t^{-1} \equiv A_1' K_1^\dagger A_1 + \cdots + A_{\lceil t \rceil - 1}' K_{\lceil t \rceil - 1}^\dagger A_{\lceil t \rceil - 1} + A_{\lceil t \rceil}' K_t^\dagger A_{\lceil t \rceil},$$

that it will be singular for the first few time periods. While the calculation of Procedure 1 doesn't depend on the invertibility of Ω_t^{-1}, the form of the target wealth W_t^* as given in Lemma 3.4.1 *is* dependent on invertibility. Since the Theorem we are about to present uses this representation of W_t^* we demand that Ω_t^{-1} become invertible by the start of some period β. In general this condition isn't very limiting. Essentially it is equivalent to demanding that the collective row space of the sequence of $m \times d$ matrices $A_1, A_2, \ldots$, have dimension d. This is typically the case for many of the target classes one would normally consider.

Given the Alternate Universality Conditions we now present the following theorem proving that Procedure 1 is universal with respect to the linear target class $\left(\mathbf{R}^d, S, A(s)\theta\right)$,

Theorem 3.4.3 *Suppose that the Alternate Universality Conditions hold. Then Procedure 1 is universal with respect to the linear target class* $\left(\mathbf{R}^d, S, A(s)\theta\right)$ *in the sense that for any* $t > \max\left\{1+\beta, \lambda_{\max}\left(\Lambda^{-1}\right)/L_\Omega^-\right\}$

$$\frac{W_t^*}{\widehat{W}_t} \le C\left[\lambda_{\min}\left(\Lambda^{-1}\right) L_\Omega^+ \lceil t\rceil + 1\right]^{d/2}.$$

Here,

$$\begin{aligned} C &= \exp\left\{2L_A\left(L_\mu + \frac{m}{2}L_K\right)^2\left(\frac{\lambda_{\max}(\Lambda^{-1})}{(L_\Omega^-)^2}\right)\right. \\ &\quad \left. +2\frac{L_A^{1/2}}{L_\Omega^-}\left(L_\mu + \frac{m}{2}L_K\right)\left\|\Lambda^{-1}\lambda\right\| + \frac{1}{2}\lambda'\Lambda^{-1}\lambda\right\}\end{aligned}$$

is a constant independent of t. As always $W_t^* = \max_{\theta\in\mathbf{R}^d} W_t(\theta)$ *denotes the hindsight optimal wealth within the target class.*

The theorem works in part because the new conditions are sufficiently strong to insure that the optimal parameters, $\theta_t^* = \arg\max_{\theta\in\Theta} W_t(\theta)$, forever stay within a compact neighborhood. The proof is as follows.

Proof. Recall wealth, $\widehat{W}_t$, of Procedure 1 is given in Lemma 3.4.2. Also from Lemma 3.4.1 we know that $W_t^* = W_0 \exp\left\{\frac{1}{2}v_t'\Omega_t^{-1}v_t\right\}$. Combining these expressions we write,

$$\frac{W_t^*}{\widehat{W}_t} = \left(|\Lambda|\left|\Omega_t^{-1}+\Lambda^{-1}\right|\right)^{1/2}\exp\left\{-\frac{1}{2}\Psi_t'\left(\Omega_t^{-1}+\Lambda^{-1}\right)\Psi_t + \frac{1}{2}v_t'\Omega_t^{-1}v_t + \frac{1}{2}\lambda'\Lambda^{-1}\lambda\right\}.$$

We bound each part of the expression in turn. First note that,

$$\left(|\Lambda|\left|\Omega_t^{-1}+\Lambda^{-1}\right|\right)^{1/2} = \left(\left|\Lambda\Omega_t^{-1}+I\right|\right)^{1/2}.$$

Since $\det A \le (\lambda_{\max}(A))^d$ for any $d\times d$ matrix A, it follows that,

$$\le \left[\lambda_{\max}\left(\Lambda\Omega_t^{-1}+I\right)\right]^{d/2}.$$

But $\lambda_{\max}\left(\Lambda\Omega_t^{-1}+I\right) \le \lambda_{\max}(\Lambda)\,\lambda_{\max}\left(\Omega_t^{-1}\right)+1 \le \lambda_{\max}(\Lambda)\,L_\Omega^+\lceil t\rceil+1 = \lambda_{\min}\left(\Lambda^{-1}\right)L_\Omega^+\lceil t\rceil + 1$ so,

$$\le \left[\lambda_{\min}\left(\Lambda^{-1}\right)L_\Omega^+\lceil t\rceil + 1\right]^{d/2}.$$

As for the exponential term, note that,

$$\begin{aligned}
& \exp\left\{-\frac{1}{2}\Psi_t'\left(\Omega_t^{-1}+\Lambda^{-1}\right)\Psi_t+\frac{1}{2}v_t'\Omega_t^{-1}v_t+\frac{1}{2}\lambda'\Lambda^{-1}\lambda\right\} \\
= & \exp\left\{-\frac{1}{2}\left(\Omega_t^{-1}v_t+\Lambda^{-1}\lambda\right)'\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\left(\Omega_t^{-1}v_t+\Lambda^{-1}\lambda\right)+\frac{1}{2}v_t'\Omega_t^{-1}v_t+\frac{1}{2}\lambda'\Lambda^{-1}\lambda\right\} \\
= & \exp\left\{\frac{1}{2}v_t'\Omega_t^{-1}\left(\Omega_t-\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\right)\Omega_t^{-1}v_t-v_t'\Omega_t^{-1}\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\Lambda^{-1}\lambda\right. \\
& \left.-\frac{1}{2}\lambda'\Lambda^{-1}\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\Lambda^{-1}\lambda+\frac{1}{2}\lambda'\Lambda^{-1}\lambda\right\}. \qquad (3.33)
\end{aligned}$$

We aim to bound each term in the exponent separately. In preparation for bounding the first term we claim that $\lambda_{\max}\left(\Lambda^{-1}\Omega_t\right)<1$ This is an immediate consequence of the assumption that $t>\max\left\{1+\beta,\lambda_{\max}\left(\Lambda^{-1}\right)/L_\Omega^-\right\}$. This allows us to write

$$\left(I+\Lambda^{-1}\Omega_t\right)^{-1}=I-\Lambda^{-1}\Omega_t+\left(\Lambda^{-1}\Omega_t\right)^2-\cdots.$$

The identity permits the derivation of the following bound,

$$\begin{aligned}
& \frac{1}{2}v_t'\Omega_t^{-1}\left(\Omega_t-\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\right)\Omega_t^{-1}v_t \\
= & \frac{1}{2}v_t'\Omega_t^{-1}\left(\Omega_t-\Omega_t\left(I+\Lambda^{-1}\Omega_t\right)^{-1}\right)\Omega_t^{-1}v_t \\
= & \frac{1}{2}v_t'\Omega_t^{-1}\left(\Omega_t\left(\Lambda^{-1}\Omega_t\right)-\Omega_t\left(\Lambda^{-1}\Omega_t\right)^2+\Omega_t\left(\Lambda^{-1}\Omega_t\right)^3-\cdots\right)\Omega_t^{-1}v_t \\
= & \frac{1}{2}v_t'\Omega_t^{-1}\left(\Omega_t\left(\Lambda^{-1}\Omega_t\right)\left(I-\Lambda^{-1}\Omega_t+\left(\Lambda^{-1}\Omega_t\right)^2-\cdots\right)\right)\Omega_t^{-1}v_t \\
= & \frac{1}{2}v_t'\Omega_t^{-1}\left(\Omega_t\Lambda^{-1}\Omega_t\left(I+\Lambda^{-1}\Omega_t\right)^{-1}\right)\Omega_t^{-1}v_t \\
= & \frac{1}{2}v_t'\Omega_t^{-1}\left(\Omega_t\Lambda^{-1}\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\right)\Omega_t^{-1}v_t \\
\leq & \frac{1}{2}\left\|\Omega_t^{-1}v_t\right\|^2\left(\lambda_{\max}\left(\Omega_t\Lambda^{-1}\right)\lambda_{\max}\left(\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\right)\right).
\end{aligned}$$

Now use the relation, $\lambda_{\max}(A)=1/\lambda_{\min}\left(A^{-1}\right)$ for positive definite matrices to show the above is bounded by,

$$\begin{aligned}
= & \frac{1}{2}\left\|\Omega_t^{-1}v_t\right\|^2\left(\frac{\lambda_{\max}(\Lambda^{-1})}{\lambda_{\min}\left(\Omega_t^{-1}\right)\lambda_{\min}\left(\Omega_t^{-1}+\Lambda^{-1}\right))}\right) \\
\leq & \frac{1}{2}\left\|\Omega_t^{-1}v_t\right\|^2\left(\frac{\lambda_{\max}(\Lambda^{-1})}{\lambda_{\min}^2\left(\Omega_t^{-1}\right)}\right) \\
\leq & \frac{1}{2}\left\|\Omega_t^{-1}v_t\right\|^2\left(\frac{\lambda_{\max}(\Lambda^{-1})}{\left(L_\Omega^-\right)^2\left(\lceil t\rceil-1\right)^2}\right). \qquad (3.34)
\end{aligned}$$

To bound $\|\Omega_t^{-1}v_t\|^2$ we note that,

$$\begin{aligned}\|\Omega_t^{-1}v_t\|^2 &= \|u_1+\cdots+u_{\lceil t\rceil-1}+u_t\|^2\\ &\le \left(\sum_{\tau\in I(t)}\|u_\tau\|\right)^2, \text{ where } I(t)=\{1,\ldots,\lceil t\rceil-1,t\}\end{aligned}$$

and,

$$\begin{aligned}\sum_{\tau\in I(t)}\|u_t\| &= (u_t\cdot u_t)^{1/2}\\ &= \sum_{\tau\in I(t)}\left(\left[\mu_t^\dagger+\frac{1}{2}\mathrm{diag}K_t^\dagger\right]' A_{\lceil t\rceil}A'_{\lceil t\rceil}\left[\mu_t^\dagger+\frac{1}{2}\mathrm{diag}K_t^\dagger\right]\right)^{1/2}\\ &\le \sum_{\tau\in I(t)}\lambda_{\max}\left(A_{\lceil t\rceil}A'_{\lceil t\rceil}\right)^{1/2}\left\|\mu_t^\dagger+\frac{1}{2}\mathrm{diag}K_t^\dagger\right\|\\ &\le L_A^{1/2}\sum_{\tau\in I(t)}\left(\left\|\mu_t^\dagger\right\|+\left\|\frac{1}{2}\mathrm{diag}K_t^\dagger\right\|\right)\\ &\le L_A^{1/2}\left(L_\mu+\frac{m}{2}L_K\right)\lceil t\rceil.\end{aligned}$$

Thus,

$$\|\Omega_t^{-1}v_t\|^2\le L_A\left(L_\mu+\frac{m}{2}L_K\right)^2\lceil t\rceil^2 \tag{3.35}$$

and upon using 3.35 with 3.34 we get,

$$\begin{aligned}\frac{1}{2}v_t'\Omega_t^{-1}\left(\Omega_t-\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\right)\Omega_t^{-1}v_t &\le \frac{1}{2}L_A\left(L_\mu+\frac{m}{2}L_K\right)^2\lceil t\rceil^2\left(\frac{\lambda_{\max}(\Lambda^{-1})}{(L_\Omega^-)^2(\lceil t\rceil-1)^2}\right)\\ &\le 2L_A\left(L_\mu+\frac{m}{2}L_K\right)^2\left(\frac{\lambda_{\max}(\Lambda^{-1})}{(L_\Omega^-)^2}\right), \qquad (3.36)\end{aligned}$$

where the last step follows because $\lceil t\rceil^2/(\lceil t\rceil-1)^2<4$ for $t>1$.

Continuing with the next term observe that,

$$\begin{aligned}&\left|v_t'\Omega_t^{-1}\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\Lambda^{-1}\lambda\right|\\ \le\ &\|\Omega_t^{-1}v_t\|\,\|\Lambda^{-1}\lambda\|\,\lambda_{\max}\left(\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\right)\\ \le\ &L_A^{1/2}\left(L_\mu+\frac{m}{2}L_K\right)\lceil t\rceil\,\|\Lambda^{-1}\lambda\|\,\lambda_{\max}\left(\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\right) \quad \text{from (3.35)}\end{aligned}$$

but,

$$\begin{aligned}\lambda_{\max}\left(\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\right) &= 1/\lambda_{\min}\left(\Omega_t^{-1}+\Lambda^{-1}\right)\\ &\le 1/\lambda_{\min}\left(\Omega_t^{-1}\right)\\ &\le \frac{1}{L_\Omega^-(\lceil t\rceil-1)}. \qquad (3.37)\end{aligned}$$

Combining this bound with the previous bound we conclude that,

$$\begin{aligned}\left|v_t'\Omega_t^{-1}\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\Lambda^{-1}\lambda\right| &\leq \frac{L_A^{1/2}\lceil t\rceil}{L_\Omega^-\left(\lceil t\rceil-1\right)}\left(L_\mu+\frac{m}{2}L_K\right)\left\|\Lambda^{-1}\lambda\right\| \\ &\leq 2\frac{L_A^{1/2}}{L_\Omega^-}\left(L_\mu+\frac{m}{2}L_K\right)\left\|\Lambda^{-1}\lambda\right\| \text{ for } t>1 \quad (3.38)\end{aligned}$$

Finally, due to the positive definiteness of $'\Lambda^{-1}\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\Lambda^{-1}$ we must conclude that $-\frac{1}{2}\lambda'\Lambda^{-1}\left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\Lambda^{-1}\lambda. < 0$. This, along with bounds (3.36) and (3.38) is sufficient to bound the exponent of (3.33) and we conclude that,

$$\frac{W_t^*}{\widehat{W}_t^{(1)}} \leq C\left[\lambda_{\min}\left(\Lambda^{-1}\right)L_\Omega^+\lceil t\rceil+1\right]^{d/2}$$

where,

$$C=\exp\left\{2L_A\left(L_\mu+\frac{m}{2}L_K\right)^2\left(\frac{\lambda_{\max}(\Lambda^{-1})}{\left(L_\Omega^-\right)^2}\right)+2\frac{L_A^{1/2}}{L_\Omega^-}\left(L_\mu+\frac{m}{2}L_K\right)\left\|\Lambda^{-1}\lambda\right\|+\frac{1}{2}\lambda'\Lambda^{-1}\lambda\right\}.$$

■

Chapter 4

Universality in Near-Continuous Time

In the previous chapter we developed continuously traded procedures, universal with respect to the linearly parameterized target class $(\mathbf{R}^d, S, A(s)\theta)$ that had the desirable property of being easily computable at any time instance t. While showing the existence of these procedures is a step in the right direction, there is still the problem of applicability. In the real world, trading isn't done in a continuous fashion, it's done discretely, so these procedures are of little use in their present form. In order to make them tradeable, we clearly have to discretize them in some way. The challenge before us then is to take discrete analogs of these continuous procedures in such a way that universality and ease of computation are preserved. Intuitively it would seem that this should be possible. After all, if we take a discrete procedure and rebalance it frequently enough, it is almost as if we were trading in continuous time. Hence, using ever increasing rates of rebalancing it would seem that we should converge to the continuous time results and therefore achieve our goals.

In this chapter we argue that a direct discrete analog of the continuously traded Procedure 1 of section 3.4.1 can be made both easily computable and universal with respect to the continuously traded linear class $(\mathbf{R}^d, S, A(s)\theta)$. We begin with some preliminaries, defining the discrete investment environment and proposed universal procedure. We then proceed with the bulk of the chapter, proving the necessary lemmas and theorems to show that our discrete procedure comes within an arbitrary small factor of the optimal growth rate. We then surmise that if rebalancing is conducted on an ever increasing schedule the resulting discrete procedure should be universal with respect to the continuously traded target class. After this is accomplished we present two implementations of this discrete procedure on market data. Finally we conclude with a discussion of the procedures computational properties, arguing that the proposed procedure is computable within a constant factor of d^2 steps (d being the dimensionality of our parameter space).

4.1 Introduction

4.1.1 Definitions

Let us consider once more investment in m stocks and cash over T periods. Taking cues from the previous chapter, we wish to consider sets of portfolio sequences defined by parameters $\theta \in \mathbf{R}^d$, side information $s \in S$, and a linearly parameterized portfolio mapping $b(\theta, s) = \left(b_0(\theta, s), \widetilde{b}(\theta, s)\right)$ where,

$$\widetilde{b}(\theta, s) = \left(b_1(\theta, s), \ldots, b_m(\theta, s)\right) = A(s)\theta$$

represents proportions of wealth in stock, $b_0(\theta, s) = 1 - \sum_{j=1}^{m} b_m(\theta, s)$ represents the proportion of wealth in cash, and $A(s)$ represents some $m \times d$ matrix dependent on the state of side information s. At the beginning of each period, the portfolio map and available side information are used to set a constant rebalanced portfolio which is then rebalanced a total of n times in the period. Thus, at the beginning of the first period we take side information s_1 to set the constant rebalanced portfolio $b_1 = b(\theta, s_1)$ having proportions of wealth in stock $\widetilde{b}_1 = A(s_1)\theta = A_1\theta$. We then take b_1 and rebalance it n times over time period $[0, 1)$. The timing of these rebalances are evenly spaced and occur at times $t \in \{0, 1/n, 2/n, \ldots, (n-1)/n\}$. At the start of the second period, we set portfolio $b_2 = b(\theta, s_2)$ with corresponding stock proportions $\widetilde{b}_2 = A(s_2)\theta = A_2\theta$ and proceed to rebalance it n times over time period $[1, 2)$. This process repeats itself for T periods until we reach our investment horizon at time $t = T$. We can see that this is nearly the same as our setting in the previous chapter except now we are rebalancing the constant rebalanced portfolios only n times in a period as opposed to the infinite number of times we used in the continuous case.

4.1.2 Wealth of Linear Classes

As before, we use $P_{t,j}$ to denote the price of stock j at time t. Given some initial wealth W_0, the wealth achieved by $b(\theta, s)$ up to times $t = k/n$, $k \in \{1, \ldots, Tn\}$ is given by,

$$\begin{aligned} W_{k/n}^{(n)}(\theta) &= W_0 \prod_{h=1}^{k} \left(b_0\left(\theta, s_{\lceil h/n \rceil}\right) + \sum_{j=1}^{m} b_j\left(\theta, s_{\lceil h/n \rceil}\right) \frac{P_{h/n,j}}{P_{(h-1)/n,j}} \right) \\ &= W_0 \prod_{h=1}^{k} \left(1 + \sum_{j=1}^{m} b_j\left(\theta, s_{\lceil h/n \rceil}\right) \left(\frac{P_{h/n,j}}{P_{(h-1)/n,j}} - 1 \right) \right). \end{aligned}$$

The first step towards getting easily computable universal procedures in this discrete time setting is finding an approximate expression for $W_{k/n}^{(n)}(\theta)$ that looks like the continuous time counterpart given in equation (3.4). To this end we define as before measures of empirical drift and covariance based on log prices $\log Z_{t,j} = P_{t,j}$. Thus define the empirical log-drift from the beginning of the most recent time period as of time $t = k/n$ by,

$$\mu_{k/n}^{\dagger} = (\mu_{k/n,1}^{\dagger}, \dots, \mu_{k/n,m}^{\dagger}) = (Z_{k/n,1} - Z_{\lceil k/n \rceil -1,1}, \dots, Z_{k/n,m} - Z_{\lceil k/n \rceil -1,m}).$$

Similarly empirical covariation from the beginning of the most recent time period as of time $t = k/n$ is defined by the $m \times m$ matrix $K_{k/n}^{\dagger(n)}$ having entries,

$$K_{k/n,i,j}^{\dagger(n)} = \sum_{h=n(\lceil k/n \rceil -1)+1}^{k} \left(Z_{h/n,i} - Z_{(h-1)/n,i}\right) \times \left(Z_{h/n,j} - Z_{(h-1)/n,j}\right).$$

In order to derive an approximation of $W_{k/n}^{(n)}(\theta)$ we will again need to assume some basic properties about our price paths. Unless noted otherwise we will always assume the following:

Minimal Path Condition

1. There exists a constant $L_P > 0$, such that,

$$\left(1 + \frac{L_P\,(1+\log n)}{\sqrt{n}}\right)^{-1} \le \frac{P_{k/n,j}}{P_{(k-1)/n,j}} \le 1 + \frac{L_P\,(1+\log n)}{\sqrt{n}}$$

for all $n \in \mathbf{N}$, $T \in \mathbf{N}$ and $k \in \{1, \dots, Tn\}$.

Given these definitions and assumptions we now present a lemma that gives an approximation of wealth $W_{k/n}^{(n)}(\theta)$ achieved by these linearly parameterized strategies considered thus far. The approximation rests on being able to bound the L_1-norm of portfolios used by a given strategy indexed by θ.

Lemma 4.1.1 *Assume the minimal path condition holds. Suppose we invest and rebalance according to* $\tilde{b}\left(\theta, s_{\lfloor k/n \rfloor +1}\right) = A_{\lfloor k/n \rfloor +1}\theta$ *at times* $t \in \{k/n : k \in \{0, \dots, Tn\}\}$. *If there is some constant* $B(n) < \frac{\sqrt{n}}{L_P(1+\log n)}$ *possibly depending on n such that* $\left\|A_{\lfloor k/n \rfloor +1}\theta\right\|_1 < B(n)$ *for all k then the wealth achieved by time k/n is,*

$$W_{k/n}^{(n)}(\theta) = W_0 \exp\left\{ \sum_{\tau \in I(k/n)} \mu_{\tau}^{\dagger\prime} A_{\lceil \tau \rceil}\theta + \frac{1}{2} \sum_{\tau \in I(k/n)} \sum_{j=1}^{m} K_{\tau,j,j}^{\dagger(n)} A_{\lceil \tau \rceil}\theta \,. - \sum_{\tau \in I(k/n)} \frac{1}{2}\theta A_{\lceil \tau \rceil}' K_{k/n,j,j}^{\dagger(n)} A_{\lceil \tau \rceil}\theta + \varepsilon_{k/n}(\theta) \right\}$$

where $I(k/n) = \{1, \ldots, \lceil k/n \rceil - 1, k/n\}$ *and* $\varepsilon_{k/n}(\theta)$ *is an* $O\left(B^3(n)\, k\,(1+\log n)^3/n^{3/2}\right)$ *remainder term bounded according to,*

$$
\begin{aligned}
\left|\varepsilon_{k/n}(\theta)\right| \;\le\; & \frac{k}{n}\left(\frac{B^3(n)\, L_P^3\,(1+\log n)^3}{3\sqrt{n}\,(1-B(n)\,L_P\,(1+\log n)/\sqrt{n})^3} + \frac{B^2(n)\, L_P^3\,(1+\log n)^3}{2\sqrt{n}}\left(1+L_P/\sqrt{n}\right)^2\right.\\
&+\frac{B^2(n)\, L_P^4\,(1+\log n)^4}{4n}\left(1+L_P\,(1+\log n)/\sqrt{n}\right)^4\\
&+\frac{B(n)\, L_P^3\,(1+\log n)^3}{2\sqrt{n}}\left(1+L_P\,(1+\log n)/\sqrt{n}\right)^2\\
&+\frac{B(n)\, L_P^4\,(1+\log n)^4}{4n}\left(1+L_P\,(1+\log n)/\sqrt{n}\right)^4\\
&\left.+\frac{B(n)\, L_P^3\,(1+\log n)^3}{3\sqrt{n}}\left(1+L_P\,(1+\log n)/\sqrt{n}\right)^3\right).
\end{aligned}
$$

Moreover, the distance between consecutive remainder terms, $\left|\varepsilon_{k/n}(\theta) - \varepsilon_{(k-1)/n}(\theta)\right|$ *is* $O\left(B^3(n)\,(1+\log n)^3/n^{3/2}\right)$.

The lemma is significant in the sense that it shows that wealth is approximately exponentially quadratic in θ. This will be important for the universality proofs of the next section.

Proof. The theorem is stated for a linear mapping $b(\theta, s) = A(s)\theta$ but in actuality the theorem holds for an arbitrary mapping as long as it satisfies the L_1-bound condition. Regardless we assume that, $\left\|\widetilde{b}\left(\theta, s_{\lfloor k/n \rfloor+1}\right)\right\|_1 < B(n) < \frac{\sqrt{n}}{L_P(1+\log n)}$ for all $k \in \{0, \ldots, Tn\}$. The wealth achieved by strategy

$$b(\theta, s) = \left(b_0(\theta, s), \widetilde{b}(\theta, s)\right) = (b_0(\theta, s), b_1(\theta, s), \ldots, b_m(\theta, s))$$

by time $t = k/n$ is,

$$
\begin{aligned}
W_{k/n}^{(n)}(\theta) &= W_0 \prod_{h=1}^{k}\left(b_0(\theta, s_{\lceil h/n \rceil}) + \sum_{j=1}^{m} b_j(\theta, s_{\lceil h/n \rceil})\frac{P_{h/n,j}}{P_{(h-1)/n,j}}\right)\\
&= W_0 \prod_{h=1}^{k}\left(1 + \sum_{j=1}^{m} b_j(\theta, s_{\lceil h/n \rceil})\left(\frac{P_{h/n,j}}{P_{(h-1)/n,j}} - 1\right)\right).
\end{aligned}
$$

To simplify exposition we henceforth define $b_{h,j} \equiv b_j(\theta, s_{\lceil h/n \rceil})$. Now set $R_{h,j}^{(n)} \equiv \left(\frac{P_{h/n,j}}{P_{(h-1)/n,j}} - 1\right)$ and continue by observing that the above is equal to,

$$= W_0 \exp\left\{\sum_{h=1}^{k} \log\left(1 + \sum_{j=1}^{m} b_{h,j} R_{h,j}^{(n)}\right)\right\}.$$

Now apply expansion $\log(1+x) = x - x^2/2 + x^3/3(1+c)^3$ (for some c between x and 0). Replacing x with $\sum_{j=1}^{m} b_{h,j} R_{h,j}^{(n)}$ we write,

$$= W_0 \exp\left\{ \sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}R_{h,j}^{(n)} - \frac{1}{2}\sum_{h=1}^{k}\left(\sum_{j=1}^{m} b_{h,j}R_{h,j}^{(n)}\right)^2 + \sum_{h=1}^{k}\frac{1}{3(1+c_h)^3}\left(\sum_{j=1}^{m} b_{h,j}R_{h,j}^{(n)}\right)^3 \right\}. \tag{4.1}$$

We work with each term in the exponent separately starting with the last. From given assumptions,

$$\begin{aligned} \left|\sum_{j=1}^{m} b_{h,j}R_{h,j}^{(n)}\right| &\leq \|b_h\|_1 \max\left|R_{h,j}^{(n)}\right| \\ &\leq \frac{B(n)\,L_P\,(1+\log n)}{\sqrt{n}}. \end{aligned}$$

Thus,

$$\begin{aligned} &\left|\sum_{h=1}^{k}\frac{1}{3(1+c_h)^3}\left(\sum_{j=1}^{m} b_{h,j}R_{h,j}^{(n)}\right)^3\right| \\ \leq\ & \sum_{h=1}^{k}\left|\frac{1}{3(1+c_h)^3}\right|\left(B(n)\,L_P\,(1+\log n)\sqrt{m}/\sqrt{n}\right)^3 \\ \leq\ & \frac{k\left(B(n)\,L_P\,(1+\log n)/\sqrt{n}\right)^3}{3\left(1-B(n)\,L_P/\sqrt{n}\right)^3} \\ =\ & \frac{k}{n}\left(\frac{B^3(n)\,L_P^3\,(1+\log n)^3}{3\sqrt{n}\left(1-B(n)\,L_P\,(1+\log n)/\sqrt{n}\right)^3}\right). \end{aligned} \tag{4.2}$$

Continuing with the middle term of (4.1), we use the expansion $x = \log(1+x) + x^2/\left(2(1+c)^2\right)$ (for some c between 0 and x) and replace x with $R_{h,j}^{(n)}$ to get,

$$\begin{aligned}\frac{1}{2}\sum_{h=1}^{k}\left(\sum_{j=1}^{m} b_{h,j}R_{h,j}^{(n)}\right)^2 &= \frac{1}{2}\sum_{h=1}^{k}\left(\sum_{j=1}^{m} b_{h,j}\log\left(1+R_{h,j}^{(n)}\right)+\sum_{j=1}^{m} b_{h,j}\frac{R_{h,j}^{(n)2}}{2\left(1+c_{h,j}\right)^2}\right)^2\\ &= \frac{1}{2}\sum_{h=1}^{k}\left(\sum_{j=1}^{m} b_{h,j}\log\left(1+R_{h,j}^{(n)}\right)\right)^2\\ &\quad+\sum_{h=1}^{k}\left(\sum_{j=1}^{m} b_{h,j}\log\left(1+R_{h,j}^{(n)}\right)\sum_{j=1}^{m} b_{h,j}\frac{R_{h,j}^{(n)2}}{2\left(1+c_{h,j}\right)^2}\right)\\ &\quad+\frac{1}{2}\sum_{h=1}^{k}\left(\sum_{j=1}^{m} b_{h,j}\frac{R_{h,j}^{(n)2}}{2\left(1+c_{h,j}\right)^2}\right)^2. \end{aligned} \tag{4.3}$$

Again, we address each of these terms separately. First note that,

$$\log\left(1+R_{h,j}^{(n)}\right) = \left(Z_{h/n,j} - Z_{(h-1)/n,j}\right).$$

Hence,

$$\begin{aligned}&\frac{1}{2}\sum_{h=1}^{k}\left(\sum_{j=1}^{m} b_{h,j}\log\left(1+R_{h,j}^{(n)}\right)\right)^2\\ &= \frac{1}{2}\sum_{h=1}^{k}\sum_{i=1}^{m}\sum_{j=1}^{m} b_{h,i}b_{h,j}\left(Z_{h/n,i}-Z_{(h-1)/n,i}\right)\left(Z_{h/n,j}-Z_{(h-1)/n,j}\right)\\ &= \frac{1}{2}\sum_{\tau\in I(k/n)}\sum_{i=1}^{m}\sum_{j=1}^{m} b_i(\theta, s_{\lceil\tau\rceil})b_j(\theta, s_{\lceil\tau\rceil})K_{\tau,i,j}^{\dagger(n)}\\ &= \frac{1}{2}\sum_{\tau\in I(k/n)} \tilde{b}'(\theta, s_{\lceil\tau\rceil})K_{\tau,j,j}^{\dagger(n)}\tilde{b}(\theta, s_{\lceil\tau\rceil}). \end{aligned} \tag{4.4}$$

Moving along, we now try to bound the second term of (4.3). It will help to note that

$$\left|\log\left(1+R_{h,j}^{(n)}\right)\right| \le L_P\left(1+\log n\right)/\sqrt{n}$$

and

$$-\left(\frac{L_P\left(1+\log n\right)}{\sqrt{n}}\right)\left(1+\frac{L_P\left(1+\log n\right)}{\sqrt{n}}\right)^{-1} \le R_{h,j}^{(n)} \le \frac{L_P\left(1+\log n\right)}{\sqrt{n}}.$$

With these bounds in mind, observe that,

$$\begin{aligned}
& \left|\sum_{h=1}^{k}\left(\sum_{j=1}^{m} b_{h,j}\log\left(1+R_{h,j}^{(n)}\right)\sum_{j=1}^{m} b_{h,j}\frac{R_{h,j}^{(n)2}}{2\left(1+c_{h,j}\right)^{2}}\right)\right| \\
\leq\ & \sum_{h=1}^{k}\sum_{i=1}^{m}\sum_{j=1}^{m} b_{h,i}b_{h,j}\left|\log\left(1+R_{h,i}^{(n)}\right)\right|\left|\frac{R_{h,j}^{(n)2}}{2\left(1+c_{h,j}\right)^{2}}\right| \\
\leq\ & \sum_{h=1}^{k}\sum_{i=1}^{m}\sum_{j=1}^{m} b_{h,i}b_{h,j}\left(\frac{L_P\left(1+\log n\right)}{\sqrt{n}}\right)\frac{L_P^2\left(1+\log n\right)^2/n}{2\left(1-\left(\frac{L_P(1+\log n)}{\sqrt{n}}\right)\left(1+\frac{L_P(1+\log n)}{\sqrt{n}}\right)^{-1}\right)^2} \\
\leq\ & k\left\|\tilde{b}\right\|_1^2\left(\frac{L_P\left(1+\log n\right)}{\sqrt{n}}\right)^3\left(1+\frac{L_P\left(1+\log n\right)}{\sqrt{n}}\right)^2/2 \\
=\ & \frac{k}{n}\left(\frac{B^2\left(n\right)L_P^3\left(1+\log n\right)^3}{2\sqrt{n}}\left(1+\frac{L_P\left(1+\log n\right)}{\sqrt{n}}\right)^2\right).
\end{aligned} \tag{4.5}$$

Similarly we bound the third term of (4.3) by

$$\begin{aligned}
& \frac{1}{2}\left|\sum_{h=1}^{k}\left(\sum_{j=1}^{m} b_{h,j}\frac{R_{h,j}^{(n)2}}{2\left(1+c_{h,j}\right)^{2}}\right)^2\right| \\
\leq\ & \sum_{h=1}^{k}\sum_{i=1}^{m}\sum_{j=1}^{m} b_{h,i}b_{h,j}\left|\frac{R_{h,i}^{(n)2}}{2\left(1+c_{h,i}\right)^{2}}\right|\left|\frac{R_{h,i}^{(n)2}}{2\left(1+c_{h,i}\right)^{2}}\right| \\
\leq\ & \sum_{h=1}^{k}\sum_{i=1}^{m}\sum_{j=1}^{m} b_{h,i}b_{h,j}\left(\frac{\left(\frac{L_P(1+\log n)}{\sqrt{n}}\right)^2}{2\left(1-\left(\frac{L_P(1+\log n)}{\sqrt{n}}\right)\left(1+\frac{L_P(1+\log n)}{\sqrt{n}}\right)^{-1}\right)^2}\right)^2 \\
\leq\ & k\left\|\tilde{b}\right\|_1^2\left(\frac{L_P\left(1+\log n\right)}{\sqrt{n}}\right)^4\left(1+\frac{L_P\left(1+\log n\right)}{\sqrt{n}}\right)^4/4 \\
=\ & \frac{k}{n}\left(\frac{B\left(n\right)^2L_P^4\left(1+\log n\right)^4}{4n}\right)\left(1+\frac{L_P\left(1+\log n\right)}{\sqrt{n}}\right)^4.
\end{aligned} \tag{4.6}$$

The remainder of the proof focuses on bounding the first term of (4.1). Using the expansion $x=\log\left(1+x\right)+x^2/2-x^3/3\left(1+c\right)^3$ (for some c between 0 and x) on each $R_{h,j}^{(n)}$, note that this term equals,

$$\begin{aligned}
\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}R_{h,j}^{(n)} =\ & \sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\log\left(1+R_{h,j}^{(n)}\right)+\frac{1}{2}\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}R_{h,j}^{(n)2} \\
& -\frac{1}{3}\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\frac{R_{h,j}^{(n)3}}{\left(1+c_{h,j}\right)^3}.
\end{aligned} \tag{4.7}$$

We examine each part of (4.7) in turn. For the first term note that,

$$\begin{aligned}
\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\log\left(1+R_{h,j}^{(n)}\right) &= \sum_{j=1}^{m}\sum_{h=1}^{k} b_{h,j}\left(Z_{h/n,i}-Z_{(h-1)/n,i}\right) \\
&= \sum_{j=1}^{m}\sum_{\tau\in I(k/n)} b_j(\theta,s_{\lceil\tau\rceil})\sum_{h=n(\lceil\tau\rceil-1)+1}^{\tau n}\left(Z_{h/n,i}-Z_{(h-1)/n,i}\right) \\
&= \sum_{j=1}^{m}\sum_{\tau\in I(k/n)} b_j(\theta,s_{\lceil\tau\rceil})\left(Z_{\tau,i}-Z_{\lceil\tau\rceil-1,i}\right) \\
&= \sum_{j=1}^{m}\sum_{\tau\in I(k/n)} b_j(\theta,s_{\lceil\tau\rceil})\mu_\tau^{\dagger} = \sum_{\tau\in I(k/n)} b'(\theta,s_{\lceil\tau\rceil})\mu_\tau^{\dagger}.
\end{aligned}$$

For the second term of (4.7) we use the expansion $x=\log(1+x)+x^2/\left(2(1+c)^2\right)$ to show that,

$$\begin{aligned}
\frac{1}{2}\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}R_{h,j}^{(n)2} &= \frac{1}{2}\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\left(\log\left(1+R_{h,j}^{(n)}\right)+\frac{R_{h,i}^{(n)2}}{2\left(1+c_{h,i}\right)^2}\right)^2 \\
&= \frac{1}{2}\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\left(\log\left(1+R_{h,j}^{(n)}\right)\right)^2 \\
&\quad+\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\log\left(1+R_{h,j}^{(n)}\right)\frac{R_{h,i}^{(n)2}}{2\left(1+c_{h,i}\right)^2} \\
&\quad+\frac{1}{2}\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\left(\frac{R_{h,i}^{(n)2}}{2\left(1+c_{h,i}\right)^2}\right)^2.
\end{aligned} \tag{4.8}$$

Observe that the first term of (4.8) is,

$$\begin{aligned}
\frac{1}{2}\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\left(\log\left(1+R_{h,j}^{(n)}\right)\right)^2 &= \frac{1}{2}\sum_{j=1}^{m}\sum_{h=1}^{k} b_{h,j}\left(Z_{h/n,i}-Z_{(h-1)/n,i}\right)^2 \\
&= \frac{1}{2}\sum_{j=1}^{m}\sum_{\tau\in I(k/n)} b_j(\theta,s_\tau).K_{\tau,j,j}^{\dagger(n)}.
\end{aligned}$$

To bound the second term of (4.8) we see that this is almost the same as (4.5) so it is almost immediate that,

$$\begin{aligned}
&\left|\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\log\left(1+R_{h,j}^{(n)}\right)\frac{R_{h,i}^{(n)2}}{2\left(1+c_{h,i}\right)^2}\right| \\
&\le \frac{k}{n}\left(\frac{B(n)\,L_P^3\left(1+\log n\right)^3}{2\sqrt{n}}\left(1+L_P\left(1+\log n\right)/\sqrt{n}\right)^2\right).
\end{aligned}$$

Similarly, we draw analogs between the third term of (4.8) and (4.6) to conclude that,

$$\left|\frac{1}{2}\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\left(\frac{R_{h,i}^{(n)2}}{2\left(1+c_{h,i}\right)^{2}}\right)^{2}\right|$$
$$\leq \frac{k}{n}\left(\frac{B(n)\,L_P^4\left(1+\log n\right)^4}{4n}\left(1+L_P\left(1+\log n\right)/\sqrt{n}\right)^4\right).$$

Using similar arguments we bound the third term of (4.7) according to,

$$\left|\frac{1}{3}\sum_{h=1}^{k}\sum_{j=1}^{m} b_{h,j}\frac{R_{h,j}^{(n)3}}{\left(1+c_{h,j}\right)^{3}}\right|$$
$$\leq \frac{k}{n}\left(\frac{B(n)\,L_P^3\left(1+\log n\right)^3}{3\sqrt{n}}\left(1+L_P\left(1+\log n\right)^3/\sqrt{n}\right)^3\right).$$

Replacing these various expressions and bounds in equation (4.1) we get the stated expressions for $W_{k/n}^{(n)}(\theta)$ and bound on $\left|\varepsilon_{k/n}(\theta)\right|$.

Finally to argue the bound on $\left|\varepsilon_{k/n}(\theta)-\varepsilon_{(k-1)/n}(\theta)\right|$ we can take each term contributing to $\varepsilon_{k/n}(\theta)$ (i.e. the LHS's of (4.2), (4.5), etc.). and subtract those that correspond to $\varepsilon_{(k-1)/n}(\theta)$. The residual terms are bounded in analogous manner to the rest of the proof. Upon doing these calculations we conclude that $\left|\varepsilon_{k/n}(\theta)-\varepsilon_{(k-1)/n}(\theta)\right|$ is $O\left(B(n)^3\left(1+\log n\right)^3/n^{3/2}\right)$. ■

4.2 Achieving Universality in Near Continuous Time

4.2.1 A Simple Adaptation

Recall that in the previous chapter we showed that the continuously traded portfolio strategy $\widehat{b}_t=\left(\widehat{b}_{t,0},\widehat{\widetilde{b}}_t\right)$,

$$\widehat{\widetilde{b}}_t=\frac{\int_{\mathbf{R}^d}A_{\lfloor t\rfloor+1}\theta W_t(\theta)\,d\pi(\theta)}{\int_{\mathbf{R}^d}W_t(\theta)\,d\pi(\theta)},\quad \pi(\theta)\sim N(\lambda,\Lambda) \tag{4.9}$$

is universal with respect to the continuously traded linearly parameterized target class $\left(\mathbf{R}^d,S,A(s)\theta\right)$. Recall also that this portfolio was shown to be easily computed at any instance t through the formula,

$$\widehat{\widetilde{b}}_t=A_{\lfloor t\rfloor+1}\Psi_t \tag{4.10}$$

where Ψ_t is computed through the relations,

$$\begin{aligned}\Psi_t &= \left(\Omega_t^{-1}+\Lambda^{-1}\right)^{-1}\left(\sum_{\tau\in I(t)} u_\tau + \Lambda^{-1}\lambda\right), \text{ with } I(t)=\{1,\ldots,\lceil t\rceil-1,t\}\\ \Omega_t^{-1} &\equiv A_1'K_1^\dagger A_1+\cdots+A_{\lceil t\rceil-1}'K_{\lceil t\rceil-1}^\dagger A_{\lceil t\rceil-1}+A_{\lceil t\rceil}'K_t^\dagger A_{\lceil t\rceil},\\ u_t &\equiv A_{\lceil t\rceil}'\left[\mu_t^\dagger+\frac{1}{2}\text{diag}K_t^\dagger\right].\end{aligned}$$

What we desire is a discrete time analog of (4.9). The most obvious choice of procedure is to rebalance wealth according to

$$\tilde{\tilde{b}}_{k/n}^{(n)}=\frac{\int_{\mathbf{R}^d} A_{\lfloor k/n\rfloor+1}\theta E_{k,n}(\theta)\,d\pi(\theta)}{\int_{\mathbf{R}^d} E_{k,n}(\theta)\,d\pi(\theta)} \tag{4.11}$$

at each rebalancing time $t=k/n$. Here,

$$\begin{aligned}E_{k,n}(\theta) &= W_0\exp\left\{\sum_{\tau\in I(k/n)}\mu_\tau^{\dagger\prime}A_{\lceil\tau\rceil}\theta+\frac{1}{2}\sum_{\tau\in I(k/n)}\sum_{j=1}^{m}K_{\tau,j,j}^{\dagger(n)}A_{\lceil\tau\rceil}\theta\right.\\ &\quad\left.-\sum_{\tau\in I(k/n)}\frac{1}{2}\theta A_{\lceil\tau\rceil}'K_{k/n,j,j}^{\dagger(n)}A_{\lceil\tau\rceil}\theta\right\}\end{aligned} \tag{4.12}$$

is just the estimate of $W_{k/n}^{(n)}(\theta)$ presented in Lemma 4.1.1 with the remainder term omitted. The reasoning behind our choice of procedure is simple. As the number of rebalances becomes large (i.e. as n gets large) the function $E_{k,n}(\theta)$ should be a good approximation to, $W_{k/n}^{(n)}(\theta)$, the wealth achieved by trading $A(s)\theta$ discretely n times each period. In turn as n becomes large $W_{k/n}^{(n)}(\theta)$ becomes a good approximation to wealth $W_{k/n}(\theta)$ achieved by trading $A(s)\theta$ continuously each period (i.e. $W_{k/n}(\theta)$ computed according to (3.4)). Because of these convergencies, $\tilde{\tilde{b}}_{k/n}^{(n)}$ is in a sense close to $\tilde{\tilde{b}}_t$ and thus we expect the wealths of each strategy to behave approximately the same.

In addition to these nice limiting properties, $\tilde{\tilde{b}}_{k/n}^{(n)}$ is easy to compute for the same reasons that $\tilde{\tilde{b}}_t$ is easy to compute. Since $E_{k,n}(\theta)$ is an exponential quadratic in θ it follows that,

$$\frac{E_{k,n}(\theta)\,d\pi(\theta)}{\int_{\mathbf{R}^d} E_{k,n}(\theta)\,d\pi(\theta)}$$

is a Gaussian measure. Thus from the form of (4.11) we see that $\tilde{\tilde{b}}_{k/n}^{(n)}$ is equivalent to the "expectation" of $A_{\lfloor k/n\rfloor+1}\theta$ under this measure. Using the properties of Gaussian expectations we quickly arrive at a simple formula for $\tilde{\tilde{b}}_{k/n}^{(n)}$. Reworking Lemma (3.4.2) it quickly follows that,

$$\overset{\asymp(n)}{b}_{k/n} = A_{\lfloor k/n \rfloor+1}\Psi_{k/n} \tag{4.13}$$

where the quantity $\Psi_{k/n}$ is now computed through the relations,

$$\begin{aligned}
\Psi_{k/n} &= \left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1}\left(\sum_{\tau\in I(k/n)} u_\tau + \Lambda^{-1}\lambda\right), \text{ with } I(k/n) = \{1,\ldots,\lceil k/n\rceil - 1, k/n\} \\
\Omega_{k/n}^{-1} &\equiv A_1' K_1^{\dagger(n)} A_1 + \cdots + A'_{\lceil k/n\rceil-1} K^{\dagger(n)}_{\lceil k/n\rceil-1} A_{\lceil k/n\rceil-1} + A'_{\lceil k/n\rceil} K^{\dagger(n)}_{k/n} A_{\lceil k/n\rceil}, \\
u_\tau &\equiv A'_{\lceil\tau\rceil}\left[\mu_\tau^\dagger + \frac{1}{2}\text{diag} K_\tau^{\dagger(n)}\right].
\end{aligned}$$

Thus $\overset{\asymp(n)}{b}_{k/n}$ has a nice closed form expression. Since we will be using this procedure frequently we define it rigorously as follows:

Procedure 2 *At time k/n invest according to $\widehat{b}^{(n)}_{k/n} = \left(\widehat{b}^{(n)}_{k/n,0}, \overset{\asymp(n)}{b}_{k/n}\right)$ where,*

$$\overset{\asymp(n)}{b}_{k/n} = \frac{\int_{\mathbf{R}^d} A_{\lfloor k/n\rfloor+1}\theta E_{k,n}(\theta)\, d\pi(\theta)}{\int_{\mathbf{R}^d} E_{k,n}(\theta)\, d\pi(\theta)} = A_{\lfloor k/n\rfloor+1}\Psi_{k/n}, \quad \pi \sim N(\lambda, \Lambda)$$

and $\widehat{b}^{(n)}_{k/n,0} = 1 - \sum_{j=1}^m \widehat{b}^{(n)}_{k/n,j}$. Here $\Psi_{k/n}$ is computed through the relations,

$$\begin{aligned}
\Psi_{k/n} &= \left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1}\left(\sum_{\tau\in I(k/n)} u_\tau + \Lambda^{-1}\lambda\right), \\
\Omega_{k/n}^{-1} &\equiv A_1' K_1^{\dagger(n)} A_1 + \cdots + A'_{\lceil k/n\rceil-1} K^{\dagger(n)}_{\lceil k/n\rceil-1} A_{\lceil k/n\rceil-1} + A'_{\lceil k/n\rceil} K^{\dagger(n)}_{k/n} A_{\lceil k/n\rceil}, \\
u_\tau &\equiv A'_{\lceil\tau\rceil}\left[\mu_\tau^\dagger + \frac{1}{2}\text{diag} K_\tau^{\dagger(n)}\right].
\end{aligned}$$

The rest of this section, which comprises the bulk of this chapter, will be solely focussed on showing that the above procedure is universal with respect to the target wealth of the corresponding *continuously traded* target class. That is, given the continuously traded target class $(\mathbf{R}^d, S, A(s)\theta)$, we will show that the above procedure traded in *discrete* time achieves to first order in the exponent the hindsight optimal wealth $W^*_{k/n} = \max_{\theta\in\mathbf{R}^d} W_{k/n}(\theta)$. We emphasize here that $W^*_{k/n}$ is achieved under continuous trading. Thus we are claiming to have a non-anticipating strategy traded in *discrete* time (i.e. Procedure 2) that tracks the wealth of the best strategy in hindsight when traded in *continuous* time. Hence, if our claim is true, we have essentially found a practical way of achieving the target wealths discussed in Chapter 3.

As groundwork for a proof of universality we will need to set some conditions on price paths, side information matrices, etc. For this reason we present the following universality conditions.

Universality Conditions (Near-Continuous Case) For all $n \in \mathbf{N}$, $T \in \mathbf{N}$ and $k \in \{1, \ldots, Tn\}$,

1. $\left(1+\frac{L_P(1+\log n)}{\sqrt{n}}\right)^{-1} \leq \frac{P_{k/n,j}}{P_{(k-1)/n,j}} \leq 1+\frac{L_P(1+\log n)}{\sqrt{n}}$ for some $L_P > 0$.

2. There exists constant $L_\mu > 0$, independent of n and k such that, $\sum_{\tau \in I(k/n)} \|\mu_\tau^{\dagger(n)}\| < L_\mu \lceil k/n \rceil$. (Recall that $I(k/n) = \{1, \ldots, \lceil k/n \rceil - 1, k/n\}$).

3. The empirical covariance matrix $K_{k/n}^{\dagger(n)}$ is positive definite. Furthermore, there exists constant $L_K > 0$ independent of n and k such that $\sum_{\tau \in I(k/n)} \lambda_{\max}(K_{k/n}^{\dagger(n)}) \leq L_K \lceil k/n \rceil$.

4. There exists a constant $L'_K > 0$ such that $|\lambda|_{\max}(K_{k/n}^{\dagger} - K_{k/n}^{\dagger(n)}) \leq \frac{L'_K}{\sqrt{n}}$.

5. For any $s \in S$, the $m \times d$ matrix $A(s)$ is of full rank and there exists positive constants $L_{A,m}$ and $L_{A,d}$ such that $\lambda_{\max}(A(s)\,A'(s)) \leq L_{A,m}$ and $\lambda_{\max}(A'(s)\,A(s)) \leq L_{A,d}$.

6. The number of periods required for $\Omega_{k/n}^{-1}$ to become invertible is at most some integer β.

7. For all $t > \beta + 1$ there exists positive constants L_Ω^- and L_Ω^+ independent of t such that $L_\Omega^-(\lceil t \rceil - 1) \leq \lambda_{\min}\left(\Omega_{k/n}^{-1}\right)$ and $\lambda_{\max}\left(\Omega_{k/n}^{-1}\right) \leq L_\Omega^+ \lceil t \rceil$.

4.2.2 Some Lemmas

The universality of Procedure 2 is difficult to prove in an overtly direct manner. In order to get to the main universality theorem it will be necessary to prove several lemmas each of which bounds the wealth of Procedure 2 in terms of the wealths of other related procedures. In preparation for these lemmas we must first prove the following lemma which bounds the size of $\Psi_{k/n}$.

Lemma 4.2.1 *Let the Universality Conditions hold. For any time k/n,*

$$\|\Psi_{k/n}\| < L_\Psi,$$

where

$$\begin{aligned} L_\Psi &= \max\left\{\frac{1}{\lambda_{\min}(\Lambda^{-1})}\left(L_{A,m}^{1/2}\left(L_\mu + \frac{\sqrt{m}}{2}L_K\right)(\beta+1) + \|\Lambda^{-1}\lambda\|\right),\right. \\ &\quad \left.\frac{\beta+2}{L_\Omega^-(\beta+1)}\,L_{A,m}^{1/2}\left(L_\mu + \frac{\sqrt{m}}{2}L_K\right) + \frac{\|\Lambda^{-1}\lambda\|}{L_\Omega^-(\beta+1)}\right\}. \end{aligned}$$

Proof. Recall from definition that,

$$\Psi_{k/n} = \left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1} \left(\sum_{\tau \in I(k/n)} u_\tau + \Lambda^{-1}\lambda \right).$$

So it follows that,

$$\begin{aligned} \left\|\Psi_{k/n}\right\| &\leq \lambda_{\max}\left(\left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1}\right) \left\| \sum_{\tau \in I(k/n)} u_\tau + \Lambda^{-1}\lambda \right\| \\ &\leq \frac{1}{\lambda_{\min}\left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)} \left(\left\| \sum_{\tau \in I(k/n)} u_\tau \right\| + \left\|\Lambda^{-1}\lambda\right\| \right). \end{aligned} \tag{4.14}$$

Now we proceed to refine this bound when $k/n \leq \beta + 1$ and $k/n > \beta + 1$. For the case of $k/n \leq \beta + 1$ note that (4.14) is bounded by,

$$\leq \frac{1}{\lambda_{\min}\left(\Lambda^{-1}\right)} \left(\left\| \sum_{\tau \in I(k/n)} u_\tau \right\| + \left\|\Lambda^{-1}\lambda\right\| \right).. \tag{4.15}$$

For the case that $k/n > \beta + 1$ we can bound (4.14) by,

$$\begin{aligned} &\leq \frac{1}{\lambda_{\min}\left(\Omega_{k/n}^{-1}\right)} \left(\left\| \sum_{\tau \in I(k/n)} u_\tau \right\| + \left\|\Lambda^{-1}\lambda\right\| \right) \\ &\leq \frac{1}{L_{\Omega}^{-}\left(\lceil k/n \rceil - 1\right)} \left(\left\| \sum_{\tau \in I(k/n)} u_\tau \right\| + \left\|\Lambda^{-1}\lambda\right\| \right) \end{aligned} \tag{4.16}$$

Now bound $\left\|\sum_{\tau \in I(k/n)} u_\tau\right\|$. Note that,

$$\begin{aligned} \left\| \sum_{\tau \in I(k/n)} u_\tau \right\| &\leq \sum_{\tau \in I(k/n)} \left\|u_\tau\right\| \\ &= \sum_{\tau \in I(k/n)} \left\| A'_{\lceil \tau \rceil} \left[\mu_\tau^{\dagger} + \frac{1}{2}\mathrm{diag} K_\tau^{\dagger(n)} \right] \right\| \\ &\leq \left(\lambda_{\max}\left(A_i A_i'\right)\right)^{1/2} \sum_{\tau \in I(k/n)} \left\| \mu_\tau^{\dagger} + \frac{1}{2}\mathrm{diag} K_\tau^{\dagger(n)} \right\| \\ &\leq L_{A,m}^{1/2} \left(L_\mu + \frac{\sqrt{m}}{2} L_K \right) \lceil k/n \rceil . \end{aligned} \tag{4.17}$$

When $k/n \leq \beta + 1$ we can use (4.17) in conjunction with (4.15) to show that,

$$\left\|\Psi_{k/n}\right\| \leq \frac{1}{\lambda_{\min}\left(\Lambda^{-1}\right)} \left(L_{A,m}^{1/2} \left(L_\mu + \frac{\sqrt{m}}{2} L_K \right) (\beta + 1) + \left\|\Lambda^{-1}\lambda\right\| \right). \tag{4.18}$$

For $k/n > \beta + 1$ we can use (4.17) in conjunction with (4.16) to show that,

$$\begin{aligned}\|\Psi_{k/n}\| &\leq \frac{1}{L_{\Omega}^{-}(\lceil k/n\rceil - 1)}\left(L_{A,m}^{1/2}\left(L_{\mu} + \frac{\sqrt{m}}{2}L_K\right)\lceil k/n\rceil + \|\Lambda^{-1}\lambda\|\right) \\ &\leq \frac{\beta+2}{L_{\Omega}^{-}(\beta+1)}L_{A,m}^{1/2}\left(L_{\mu} + \frac{\sqrt{m}}{2}L_K\right) + \frac{\|\Lambda^{-1}\lambda\|}{L_{\Omega}^{-}(\beta+1)}. \end{aligned} \tag{4.19}$$

The lemma is proven upon setting L_{Ψ} to the maximum of (4.18) and (4.19). ∎

Before proceeding further, it will serve us well to define the Gaussian measure,

$$dG_{k,n}(\theta) \equiv \frac{E_{k,n}(\theta)\, d\pi(\theta)}{\int_{\mathbf{R}^d} E_{k,n}(\theta)\, d\pi(\theta)}, .\pi \sim N(\lambda, \Lambda). \tag{4.20}$$

This measure will keep popping up in many of the lemmas and proofs presented hereafter. Having it around will greatly simplify some of the exposition. It will also be important to note at various times that the mean and covariance matrix associated with this measure are respectively,

$$\Psi_{k/n} = \left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1}\left(\sum_{\tau\in I(k/n)} u_{\tau} + \Lambda^{-1}\lambda\right)$$

and

$$\left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1}.$$

This is easily proven by revisiting the proof of Lemma 3.4.2.

Recall that our strategy for proving the universality of Procedure 2 will be to compare its wealth to that of several intermediary procedures. The first intermediary procedure we wish to consider is as follows:

Procedure 3 *Invest according to* $\widehat{b}_{k/n}^{\dagger(n)} = \left(\widehat{b}_{k/n,0}^{\dagger(n)}, \vec{b}_{k/n}^{\dagger(n)}\right)$ *where,*

$$\vec{b}_{k/n}^{\dagger(n)} = \frac{\int_{\Theta_{k,n,\varepsilon}} A_{\lfloor k/n\rfloor+1}\theta E_{k,n}(\theta)\, d\pi(\theta)}{\int_{\mathbf{R}^d} E_{k,n}(\theta)\, d\pi(\theta)} = \int_{\Theta_{k,n,\varepsilon}} A_{\lfloor k/n\rfloor+1}\theta\, dG_{k,n}(\theta)$$

and $\widehat{b}_{k/n,0}^{\dagger(n)} = 1 - \sum_{j=1}^{m}\widehat{b}_{k/n,j}^{\dagger(n)}$. *Here* $\Theta_{k,n,\varepsilon}$ *is defined for some chosen* $\varepsilon > 0$ *as,*

$$\Theta_{k,n,\varepsilon} \equiv \left\{\theta : \|\theta - \Psi_{k/n}\| \leq \frac{n^{\varepsilon/3}\log n}{4\sqrt{m}L_{A,d}^{1/2}L_P} - L_{\Psi}\right\}, \quad \textit{where } L_{\Psi} \textit{ is from Lemma 4.2.1.}$$

We see that Procedure 3 is similar to Procedure 2 with a key exception. In Procedure 2 the stock components are calculated by integrating $A_{\lfloor k/n \rfloor+1}\theta$ over all of $\mathbf{R}^d$. However in Procedure 3 we are now restricting the integration to a closed ball $\Theta_{k,n,\varepsilon}$ centered on the measure mean $\Psi_{k/n}$. The radius of $\Theta_{k,n,\varepsilon}$ should seem rather arbitrary at this point. The reasons for defining it as such will become clearer as we prove subsequent results.

Lemma 4.2.2 *Suppose the Universality Conditions are satisfied. Let $\widehat{W}_T^{(n)}$ be the wealth of Procedure 2 after T time periods and let $\widehat{W}_T^{\dagger(n)}$ be the wealth of Procedure 3 after T time periods. Then for sufficiently large n satisfying $n > \exp\left\{24\sqrt{m}L_{A,d}^{1/2}L_P L_\Psi\right\}$ and $\varepsilon > 0$ of Procedure 3, both wealths are positive and there exist constants C and α depending on ε such that,*

$$\frac{\widehat{W}_T^{\dagger(n)}}{\widehat{W}_T^{(n)}} \leq \exp\left\{T\epsilon_n\right\},$$

where,

$$\epsilon_n = Cn\exp\left\{-\alpha n^{\varepsilon/3}\log n\right\}.$$

Proof. Recall that

$$\Theta_{k,n,\varepsilon} \equiv \left\{\theta : \left\|\theta - \Psi_{k/n}\right\| \leq \frac{n^{\varepsilon/3}\log n}{4\sqrt{m}L_{A,d}^{1/2}L_P} - L_\Psi\right\}.$$

The condition $n > \exp\left\{24\sqrt{m}L_{A,d}^{1/2}L_P L_\Psi\right\}$ insures that the radius of $\Theta_{k,n,\varepsilon}$ is positive for each k. If this were not so, Procedure 3 would be ill-defined.

From basic principles we write the wealth of Procedure 2 and Procedure 3 respectively as,

$$\widehat{W}_T^{(n)} = W_0 \prod_{k=1}^{Tn} \widehat{b}_{(k-1)/n}^{(n)\prime} X_{k/n}, \qquad \widehat{W}_T^{\dagger(n)} = W_0 \prod_{k=1}^{Tn} \widehat{b}_{(k-1)/n}^{\dagger(n)\prime} X_{k/n}$$

where $X_{k/n}$ denotes the vector of wealth relatives,

$$\begin{aligned} X_{k/n} &= \left(X_{k/n,0}, X_{k/n,1}, \ldots, X_{k/n,m}\right) \\ &= \left(1, \frac{P_{k/n,1}}{P_{(k-1)/n,1}}, \ldots, \frac{P_{k/n,m}}{P_{(k-1)/n,m}}\right). \end{aligned}$$

We now wish to show that $\widehat{W}_T^{(n)}$ is positive. To see this note that,

$$\begin{aligned}\widehat{W}_T^{(n)} &= W_0 \prod_{k=1}^{Tn} \widehat{b}^{(n)'}_{(k-1)/n} X_{k/n} \\ & W_0 \prod_{k=1}^{Tn} \left(1 + \widetilde{\widetilde{b}}^{(n)'}_{(k-1)/n} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)\right),\end{aligned}$$

where,

$$\widetilde{X}_{k/n} = \left(X_{k/n,1}, \ldots, X_{k/n,m}\right),$$

and $\mathbf{1}$ is a vector of m 1's. Now recall that $\widetilde{\widetilde{b}}^{(n)'}_{(k-1)/n} = A_{\lfloor (k-1)/n \rfloor + 1} \Psi_{(k-1)/n}$ so,

$$\widehat{W}_T^{(n)} = W_0 \prod_{k=1}^{Tn} \left(1 + A_{\lfloor (k-1)/n \rfloor + 1} \Psi_{(k-1)/n} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)\right). \tag{4.21}$$

Examine each of these factors. Note that,

$$\begin{aligned}1 + A_{\lfloor (k-1)/n \rfloor + 1} \Psi_{(k-1)/n} \left(\widetilde{X}_{k/n} - \mathbf{1}\right) &\geq 1 - \left\| A_{\lfloor (k-1)/n \rfloor + 1} \Psi_{(k-1)/n} \left(\widetilde{X}_{k/n} - \mathbf{1}\right) \right\| \\ &\geq 1 - \left\| A_{\lfloor (k-1)/n \rfloor + 1} \Psi_{(k-1)/n} \right\| \left\| \widetilde{X}_{k/n} - \mathbf{1} \right\| \\ &\geq 1 - \left(L_{A,d}^{1/2} L_\Psi\right) \left(\frac{\sqrt{m} L_P (1 + \log n)}{\sqrt{n}}\right). \quad (4.22)\end{aligned}$$

But from the assumption of the lemma

$$\frac{\sqrt{n}}{1 + \log n} > \frac{\sqrt{n}}{1 + 4n^{1/4}} > \frac{n^{1/4}}{5} > \frac{1}{5} \exp\left\{6\sqrt{m} L_{A,d}^{1/2} L_P L_\Psi\right\} > \frac{6}{5} \sqrt{m} L_{A,d}^{1/2} L_P L_\Psi.$$

Substituting this into (4.22) we write,

$$1 + A_{\lfloor (k-1)/n \rfloor + 1} \Psi_{(k-1)/n} \left(\widetilde{X}_{k/n} - \mathbf{1}\right) \geq 1 - \frac{5}{6} > 0. \tag{4.23}$$

Thus the factors contributing to $\widehat{W}_T^{(n)}$ as written in (4.21) are all positive. So $\widehat{W}_T^{(n)}$ is positive and the first statement of the lemma is proven.

Now establish that $\widehat{W}_T^{\dagger(n)}$ is positive. To see this note that,

$$\begin{aligned}\widehat{W}_T^{\dagger(n)} &= W_0 \prod_{k=1}^{Tn} \widehat{b}^{\dagger(n)'}_{(k-1)/n} X_{k/n} \\ & W_0 \prod_{k=1}^{Tn} \left(1 + \widetilde{\widetilde{b}}^{\dagger(n)'}_{(k-1)/n} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)\right),\end{aligned}$$

where,

$$\widetilde{X}_{k/n} = \left(X_{k/n,1}, \cdots, X_{k/n,m}\right),$$

and $\mathbf{1}$ is a vector of m 1's. Continuing, since $\vec{b}^{\dagger(n)}_{(k-1)/n} = \int_{\Theta_{k-1,n,\epsilon}} A_{\lfloor (k-1)/n \rfloor +1}\theta dG_{k-1,n}(\theta)$,

$$\begin{aligned}
\widehat{W}^{\dagger(n)}_T &= W_0 \prod_{k=1}^{Tn} \left(1 + \left(\int_{\Theta_{k-1,n,\epsilon}} A_{\lfloor (k-1)/n \rfloor +1}\theta dG_{k-1,n}(\theta)\right)' \left(\widetilde{X}_{k/n} - \mathbf{1}\right)\right) && (4.24)\\
&= W_0 \prod_{k=1}^{Tn} \left(1 + \int_{\Theta_{k-1,n,\epsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor +1}\theta dG_{k-1,n}(\theta)\right) && (4.25)\\
&= W_0 \prod_{k=1}^{Tn} \left(1 - \int_{\Theta_{k-1,n,\epsilon}} dG_{k-1,n}(\theta)\right.\\
&\quad \left. + \int_{\Theta_{k-1,n,\epsilon}} 1 + \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor +1}\theta dG_{k-1,n}(\theta)\right). && (4.26)
\end{aligned}$$

The quantity $1 - \int_{\Theta_{k,n,\epsilon}} dG_{k-1,n}(\theta)$ is certainly positive. Consider the second integral in the factors of the above product. Note that the integrand of this integral is bounded according to,

$$\begin{aligned}
1 + \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor +1}\theta &\geq 1 - \left\|\widetilde{X}_{k/n} - \mathbf{1}\right\| \left\|A_{\lfloor (k-1)/n \rfloor +1}\theta\right\| \\
&\geq 1 - \frac{\sqrt{m}L_p(1 + \log n)}{\sqrt{n}} L_{A,d}^{1/2} \|\theta\|.
\end{aligned}$$

But since $\theta \in \Theta_{k,n,\varepsilon}$ it follows that $\|\theta\| \leq \left(\frac{n^{\varepsilon/3} \log n}{4\sqrt{m}L_{A,d}^{1/2} L_P}\right)$ and therefore,

$$\begin{aligned}
1 + \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor +1}\theta &\geq 1 - \frac{n^{\varepsilon/3}(1 + \log n) \log n}{4\sqrt{n}} \\
&> \frac{1}{4} \text{ for sufficiently small } \varepsilon.
\end{aligned}$$

Thus the integral $\int_{\Theta_{k-1,n,\epsilon}} 1 + \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor +1}\theta dG_{(k-1),n}(\theta)$ is positive and it must follow that $\widehat{W}^{\dagger(n)}_T$ is positive also.

Now we will establish the upper bound of $\widehat{W}^{\dagger(n)}_T$ in terms of $\widehat{W}^{(n)}_T$. Let us return to (4.26) and write,

$$\begin{aligned}
\widehat{W}^{\dagger(n)}_T &= W_0 \prod_{k=1}^{Tn} \left(\left(1 - \int_{\Theta_{k-1,n,\epsilon}} dG_{(k-1),n}(\theta)\right)\right.\\
&\quad \left. + \int_{\Theta_{k-1,n,\epsilon}} 1 + \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor +1}\theta dG_{(k-1),n}(\theta)\right).
\end{aligned}$$

Let us bound the quantity $\left(1 - \int_{\Theta_{k-1,n,\varepsilon}} dG_{(k-1),n}(\theta)\right)$. Recall that,

$$G_{(k-1),n} \sim N\left(\Psi_{(k-1)/n}, \left(\Omega^{-1}_{(k-1)/n} + \Lambda^{-1}\right)^{-1}\right).$$

Suppose we standardize θ through the change of variable,

$$z_{k-1}(\theta) = \left(\Omega^{-1}_{(k-1)/n} + \Lambda^{-1}\right)^{1/2} \left(\theta - \Psi_{(k-1)/n}\right).$$

Let $\mathcal{Z}_{k-1,n}$ be the image set of $\Theta_{k-1,n,\varepsilon}$ under this mapping, i.e.,

$$\mathcal{Z}_{k-1,n} = z_{k-1}\left(\Theta_{k-1,n,\varepsilon}\right).$$

If $\phi(\cdot)$ is the density of the standard d-dimensional normal, it must follow that,

$$\int_{\Theta_{k-1,n,\varepsilon}} dG_{(k-1),n}(\theta) = \int_{\mathcal{Z}_{k-1,n}} \phi(z_{k-1})\, dz_{k-1}.$$

Now further restrict the integration to subset,

$$\mathcal{Z}_n^{\dagger} = \left\{ z : \|z\| \le \lambda_{\min}\left(\Lambda^{-1/2}\right) \left(\frac{n^{\varepsilon/3} \log n}{4\sqrt{m} L_{A,d}^{1/2} L_P} - L_\Psi \right) \right\}.$$

To see that this is a subset of $\mathcal{Z}_{k-1,n}$, note that for $z \in \mathcal{Z}_n^{\dagger}$,

$$\begin{aligned} \|z\| &\le \lambda_{\min}\left(\Lambda^{1/2}\right) \left(\frac{n^{\varepsilon/3} \log n}{4\sqrt{m} L_{A,d}^{1/2} L_P} - L_\Psi \right) \\ &\le \lambda_{\min}\left(\left(\Omega^{-1}_{(k-1)/n} + \Lambda^{-1}\right)^{1/2} \right) \left(\frac{n^{\varepsilon/3} \log n}{4\sqrt{m} L_{A,d}^{1/2} L_P} - L_\Psi \right). \end{aligned} \tag{4.27}$$

Now write

$$\theta_{k-1} = \left(\Omega^{-1}_{(k-1)/n} + \Lambda^{-1}\right)^{-1/2} z + \Psi_{(k-1)/n}.$$

Inverting we get,

$$z = \left(\Omega^{-1}_{(k-1)/n} + \Lambda^{-1}\right)^{1/2} \left(\theta_{k-1} - \Psi_{(k-1)/n}\right),$$

which implies,

$$\|z\| \ge \lambda_{\min}\left(\left(\Omega^{-1}_{(k-1)/n} + \Lambda^{-1}\right)^{1/2} \right) \left\|\theta_{k-1} - \Psi_{(k-1)/n}\right\|. \tag{4.28}$$

Together (4.27) and (4.28) imply that,

$$\left\|\theta_{k-1} - \Psi_{(k-1)/n}\right\| \leq \left(\frac{n^{\varepsilon/3}\log n}{4\sqrt{m}L_{A,d}^{1/2}L_P} - L_\Psi\right),$$

which in turn implies that the θ_{k-1} is in $\Theta_{k-1,n,\varepsilon}$ and hence z is in $\mathcal{Z}_{k-1,n}$. Thus we conclude that $\mathcal{Z}_n^\dagger$ is a subset of $\mathcal{Z}_{k-1,n}$ for any k. This allows us to write,

$$\int_{\mathcal{Z}_{k-1,n}} \phi(z)\,dz \geq \int_{\mathcal{Z}_n^\dagger} \phi(z)\,dz.$$

At this point we note that we are integrating a standard normal density over a sphere growing at rate $O\left(n^{\varepsilon/3}\log n\right)$. Using the tail properties of normal densities we know that for sufficiently large n we can set positive constants C_1 and α_1 such that,

$$\int_{\mathcal{Z}_n^\dagger} \phi(z_i)\,dz_i \geq 1 - C_1\exp\left\{-\alpha_1 n^{\varepsilon/3}\log n\right\}.$$

Thus we conclude that,

$$\int_{\Theta_{k-1,n,\varepsilon}} dG_{(k-1),n}(\theta) \geq 1 - C_1\exp\left\{-\alpha_1 n^{\varepsilon/3}\log n\right\}. \tag{4.29}$$

In light of (4.29) we return to (4.26) and write,

$$\begin{aligned}
\widehat{W}_T^{\dagger(n)} &\leq W_0\prod_{k=1}^{Tn}\Big(C_1\exp\left\{-\alpha_1 n^{\varepsilon/3}\log n\right\} \\
&\quad + \int_{\Theta_{k-1,n,\varepsilon}} 1 + \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\theta dG_{(k-1),n}(\theta)\Big) \\
&= W_0\prod_{k=1}^{Tn}\Big(C_1\exp\left\{-\alpha_1 n^{\varepsilon/3}\log n\right\} \\
&\quad + \int_{\mathbf{R}^d} 1 + \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\theta dG_{(k-1),n}(\theta) \\
&\quad - \int_{\Theta_{k-1,n,\varepsilon}^C} 1 + \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\theta dG_{(k-1),n}(\theta)\Big).
\end{aligned} \tag{4.30}$$

We now endeavor to show that there exist constants C_2 and α_2 such that,

$$\left|\int_{\Theta_{k-1,n,\varepsilon}^C} 1 + \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\theta dG_{(k-1),n}(\theta)\right| \leq C_2\exp\left\{-\alpha_2 n^{\varepsilon/3}\log n\right\}.$$

Note that,

$$\left|\int_{\Theta^C_{k-1,n,\epsilon}} 1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\theta dG_{(k-1),n}(\theta)\right|$$
$$\leq \left|\int_{\Theta^C_{k-1,n,\epsilon}} \left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\left(\theta-\Psi_{(k-1)/n}\right) dG_{(k-1),n}(\theta)\right|$$
$$+\left|\int_{\Theta^C_{k-1,n,\epsilon}} 1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\Psi_{(k-1)/n} dG_{(k-1),n}(\theta)\right|. \quad (4.31)$$

We bound each term of (4.31) in turn. Taking the first term, note that,

$$\left|\int_{\Theta^C_{k-1,n,\epsilon}} \left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\left(\theta-\Psi_{(k-1)/n}\right) dG_{(k-1),n}(\theta)\right|$$
$$= \left|\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\int_{\Theta^C_{k-1,n,\epsilon}} \left(\theta-\Psi_{(k-1)/n}\right) dG_{(k-1),n}(\theta)\right|.$$

I claim that for any ball B centered on Ψ_{i-1} (of which $\Theta_{k-1,n,\epsilon}$ is one) ,

$$\int_B \left(\theta-\Psi_{(k-1)/n}\right) dG_{(k-1),n}(\theta)=0.$$

This is due to the symmetries of the Gaussian density around its mean $\Psi_{(k-1)/n}$ and the fact that we are integrating over a region centered on $\Psi_{(k-1)/n}$. For each element $\theta_1 \in B$ there is another element $\theta_2 \in B$ which is equidistant to $\Psi_{(k-1)/n}$ and on the line extending through θ_1 and $\Psi_{(k-1)/n}$. So $\left(\theta_1-\Psi_{(k-1)/n}\right) = -\left(\theta_2-\Psi_{(k-1)/n}\right)$ and $\left.\frac{dG_{(k-1),n}(\theta)}{d\theta}\right|_{\theta=\theta_1} = \left.\frac{dG_{(k-1),n}(\theta)}{d\theta}\right|_{\theta=\theta_2}$. Thus the contribution of each θ to the integral is exactly offset by another, causing the integral to vanish. Additionally, we know that in the limit,

$$\int_{\mathbf{R}^d} \left(\theta-\Psi_{(k-1)/n}\right) dG_{(k-1),n}(\theta)=0.$$

Thus when integrating over the complement B^C,

$$\int_{B^C} \left(\theta-\Psi_{(k-1)/n}\right) dG_{(k-1),n}(\theta)=0.$$

Therefore we conclude that,

$$\left|\int_{\Theta^C_{k-1,n,\epsilon}} \left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\left(\theta-\Psi_{(k-1)/n}\right) dG_{(k-1),n}(\theta)\right| = 0. \quad (4.32)$$

As for the second term of (4.31) note that,

$$\left|\int_{\Theta^C_{k-1,n,\epsilon}} 1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\Psi_{(k-1)/n} dG_{(k-1),n}(\theta)\right|$$
$$= \left|1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\Psi_{(k-1)/n}\right|\left|\int_{\Theta^C_{k-1,n,\epsilon}} dG_{(k-1),n}(\theta)\right|.$$

The first factor has already been shown to be uniformly bounded over k and n (see (4.23). The other factor, $\left|\int_{\Theta^C_{k-1,n,\epsilon}} dG_{(k-1),n}(\theta)\right|$ is bounded according to (4.29). Using these two observations we conclude that there exist constants C_2 and α_2 such that,

$$\begin{aligned}&\left|\int_{\Theta^C_{k-1,n,\epsilon}} 1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor(k-1)/n\rfloor+1}\Psi_{(k-1)/n} dG_{(k-1),n}(\theta)\right| \\ \leq\ & C_2\exp\left\{-\alpha_2 n^{\varepsilon/3}\log n\right\}. \end{aligned} \tag{4.33}$$

Substituting (4.32) and (4.33) back into (4.31) we conclude that,

$$\left|\int_{\Theta^C_{k-1,n,\epsilon}} 1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor(k-1)/n\rfloor+1}\theta dG_{(k-1),n}(\theta)\right| \leq C_2\exp\left\{-\alpha_2 n^{\varepsilon/3}\log n\right\}.$$

In turn putting this result into (4.30) yields,

$$\begin{aligned}\widehat{W}^{\dagger(n)}_T \leq\ & W_0\prod_{k=1}^{Tn}\left(C_1\exp\left\{-\alpha_1 n^{\varepsilon/3}\log n\right\}+C_2\exp\left\{-\alpha_2 n^{\varepsilon/3}\log n\right\}\right.\\ &\left.+\int_{\mathbf{R}^d} 1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor(k-1)/n\rfloor+1}\theta dG_{(k-1),n}(\theta)\right).\end{aligned}$$

Setting appropriately large C and α we can combine the first two terms into one and write,

$$\begin{aligned}\widehat{W}^{\dagger(n)}_T \leq\ & W_0\prod_{k=1}^{Tn}\left(C\exp\left\{-\alpha n^{\varepsilon/3}\log n\right\}\right.\\ &\left.+\int_{\mathbf{R}^d} 1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor(k-1)/n\rfloor+1}\theta dG_{(k-1),n}(\theta)\right).\end{aligned}$$

Now note that,

$$\int_{\mathbf{R}^d} 1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor(k-1)/n\rfloor+1}\theta dG_{(k-1),n}(\theta) = 1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor(k-1)/n\rfloor+1}\Psi_{(k-1)/n}.$$

So we write,

$$\begin{aligned}\widehat{W}^{\dagger(n)}_T \leq\ & W_0\prod_{k=1}^{Tn}\left(C\exp\left\{-\alpha n^{\varepsilon/3}\log n\right\}+1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor(k-1)/n\rfloor+1}\Psi_{(k-1)/n}\right.\\ =\ & W_0\prod_{k=1}^{Tn}\left(1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor(k-1)/n\rfloor+1}\Psi_{(k-1)/n}\right)\times\\ &\left(C\exp\left\{-\alpha n^{\varepsilon/3}\log n\right\}\left(1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor(k-1)/n\rfloor+1}\Psi_{(k-1)/n}\right)^{-1}+1\right).\end{aligned}$$

But from (4.23),

$$\left(1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\Psi_{(k-1)/n}\right)^{-1}\leq 6$$

so this factor can be absorbed into the C after increasing C slightly. Thus we write,

$$\begin{aligned}\widehat{W}_T^{\dagger(n)} &\leq W_0\prod_{k=1}^{Tn}\left(1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\Psi_{(k-1)/n}\right)\times \\ &\quad \left(C\exp\left\{-\alpha n^{\varepsilon/3}\log n\right\}+1\right). \end{aligned} \tag{4.34}$$

But now notice that,

$$\widehat{W}_T^{(n)} = W_0\prod_{k=1}^{Tn}\left(1+\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\Psi_{(k-1)/n}\right) \tag{4.35}$$

so substituting (4.35) into (4.34) we can write,

$$\begin{aligned}\widehat{W}_T^{\dagger(n)} &\leq \widehat{W}_T^{(n)}\prod_{k=1}^{Tn}\left(C\exp\left\{-\alpha n^{\varepsilon/3}\log n\right\}+1\right)\\ &= \widehat{W}_T^{(n)}\left(C\exp\left\{-\alpha n^{\varepsilon/3}\log n\right\}+1\right)^{Tn}\\ &= \widehat{W}_T^{(n)}\exp\left\{Tn\log\left(C\exp\left\{-\alpha n^{\varepsilon/3}\log n\right\}+1\right)\right\}\\ &\leq \widehat{W}_T^{(n)}\exp\left\{CTn\exp\left\{-\alpha n^{\varepsilon/3}\log n\right\}\right\}.\end{aligned}$$

Thus the lemma is proven. ■

The above lemma links the wealth of Procedure 2 to that of Procedure 3. In turn we will now link the wealth of Procedure 3 to yet another procedure. Consider the following.

Procedure 4 *At time k/n invest according to $\widehat{b}_{k/n}^{\dagger\dagger(n)} = \left(\widehat{b}_{k/n,0}^{\dagger\dagger(n)}, \widetilde{\widehat{b}}_{k/n}^{\dagger\dagger(n)}\right)$ where,*

$$\widetilde{\widehat{b}}_{k/n}^{\dagger\dagger(n)} = \frac{\int_{\Theta_{n,\varepsilon}} A_{\lfloor k/n\rfloor+1}\theta E_{k,n}(\theta)\,d\pi(\theta)}{\int_{\Theta_{n,\varepsilon}} E_{k,n}(\theta)\,d\pi(\theta)} = \frac{\int_{\Theta_{n,\varepsilon}} A_{\lfloor k/n\rfloor+1}\theta dG_{k,n}(\theta)}{\int_{\Theta_{n,\varepsilon}} dG_{k,n}(\theta)},$$

and $\widehat{b}_{k/n,0}^{\dagger\dagger(n)} = 1-\sum_{j=1}^{m}\widehat{b}_{k/n,j}^{\dagger\dagger(n)}$. Here $\Theta_{n,\varepsilon}$ is defined for some chosen $\varepsilon>0$ as,

$$\Theta_{n,\varepsilon}\equiv\left\{\theta:\|\theta\|\leq\frac{n^{\varepsilon/3}\log n}{4\sqrt{m}L_{A,d}^{1/2}L_P}\right\},\quad \textit{where } L_\Psi \textit{ is from Lemma 4.2.1.}$$

Procedure 4 is almost identical to Procedure 3. The major change is that the domain of integration for $\widetilde{b}^{\dagger\dagger(n)}_{k/n}$ is now the origin centered ball $\Theta_{n,\varepsilon}$ as opposed to the mean (i.e. $\Psi_{k/n}$) centered ball $\Theta_{k,n,\varepsilon}$. The important property of $\Theta_{n,\varepsilon}$ is that it doesn't change with k. This will be crucial in a later proof where we will need to take advantage of a collapsing telescoping product over times k/n, $k \in \{1, \ldots Tn\}$ involving integrals over $\Theta_{n,\varepsilon}$. For the time being we present the following lemma which links the wealth of Procedure 3 to that of Procedure 4.

Lemma 4.2.3 *Suppose the Universality Conditions are satisfied. Let $\widehat{W}^{\dagger(n)}_T$ be the wealth of Procedure 3 after T time periods and let $\widehat{W}^{\dagger\dagger(n)}_T$ be the wealth of Procedure 4 after T time periods. Assume that the ε's associated with both procedures are the same. Then for $n > \exp\left\{24\sqrt{m}L^{1/2}_{A,d}L_P L_\Psi\right\}$, both wealths are positive and there exist constants C, C' and α depending on ε such that,*

$$\frac{\widehat{W}^{\dagger\dagger(n)}_T}{\widehat{W}^{\dagger(n)}_T} \le C' \exp\{T\epsilon_n\},$$

where,

$$\epsilon_n = Cn \exp\left\{-\alpha n^{\varepsilon/3} \log n\right\}.$$

Proof. We have previously shown in Lemma 4.2.2 that $\widehat{W}^{\dagger(n)}_T$ is positive. To prove that $\widehat{W}^{\dagger\dagger(n)}_T$ is positive, we simply rework the $\widehat{W}^{\dagger(n)}_T$ proof for $\widehat{W}^{\dagger\dagger(n)}_T$. Since the rework is essentially the same as the original proof, we omit it and proceed with the proving the wealth bound. By definition,

$$\widehat{W}^{\dagger\dagger(n)}_T = W_0 \prod_{k=1}^{Tn} \widehat{b}^{\dagger\dagger(n)'}_{(k-1)/n} X_{k/n},$$

where $X_{k/n}$ denotes the vector of wealth relatives,

$$\begin{aligned} X_{k/n} &= \left(X_{k/n,0}, X_{k/n,1}, \ldots, X_{k/n,m}\right) \\ &= \left(1, \frac{P_{k/n,1}}{P_{(k-1)/n,1}}, \ldots, \frac{P_{k/n,m}}{P_{(k-1)/n,m}}\right). \end{aligned}$$

Thus we write,

$$\begin{aligned} \widehat{W}^{\dagger\dagger(n)}_T &= W_0 \prod_{k=1}^{Tn} \widehat{b}^{\dagger\dagger(n)'}_{(k-1)/n} X_{k/n} \\ &= W_0 \prod_{k=1}^{Tn} \left(1 + \widetilde{b}^{\dagger\dagger(n)'}_{(k-1)/n}\left(\widetilde{X}_{k/n} - 1\right)\right). \end{aligned}$$

Here,

$$\widetilde{X}_{k/n} = \left(X_{k/n,1}, \dots, X_{k/n,m}\right),$$

and $\mathbf{1}$ is a vector of m 1's. Now recall that,

$$\widetilde{b}^{\dagger\dagger(n)'}_{k/n} = \frac{\int_{\Theta_{n,\varepsilon}} A_{\lfloor k/n \rfloor + 1}\theta dG_{k,n}(\theta)}{\int_{\Theta_{n,\varepsilon}} dG_{k,n}(\theta)}$$

so,

$$\begin{aligned}
\widehat{W}^{\dagger\dagger(n)}_{T} &= W_0 \prod_{k=1}^{Tn} \left(1 + \left(\frac{\int_{\Theta_{n,\varepsilon}} A_{\lfloor (k-1)/n \rfloor + 1}\theta dG_{k-1,n}(\theta)}{\int_{\Theta_{n,\varepsilon}} dG_{k-1,n}(\theta)}\right)' \left(\widetilde{X}_{k/n} - \mathbf{1}\right)\right) \\
&= W_0 \prod_{k=1}^{Tn} \left(1 + \frac{\int_{\Theta_{n,\varepsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1}\theta dG_{k-1,n}(\theta)}{\int_{\Theta_{n,\varepsilon}} dG_{k-1,n}(\theta)}\right) \\
&\le W_0 \prod_{k=1}^{Tn} \left(\frac{1}{\int_{\Theta_{n,\varepsilon}} dG_{k-1,n}(\theta)} + \frac{\int_{\Theta_{n,\varepsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1}\theta dG_{k-1,n}(\theta)}{\int_{\Theta_{n,\varepsilon}} dG_{k-1,n}(\theta)}\right) \\
&= W_0 \prod_{k=1}^{Tn} \left(\int_{\Theta_{n,\varepsilon}} dG_{k-1,n}(\theta)\right)^{-1} \times \\
&\quad \prod_{k=1}^{Tn} \left(1 + \int_{\Theta_{n,\varepsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1}\theta dG_{k-1,n}(\theta)\right). \qquad (4.36)
\end{aligned}$$

It is evident from the definitions of $\Theta_{n,\varepsilon}$ and $\Theta_{k,n,\varepsilon}$ that $\Theta_{n,\varepsilon} \supseteq \Theta_{k,n,\varepsilon}$ for any k. Therefore,

$$\int_{\Theta_{n,\varepsilon}} dG_{k-1,n}(\theta) \ge \int_{\Theta_{k-1,n,\varepsilon}} dG_{k-1,n}(\theta).$$

Furthermore we know from equation (4.29) and its development that there exists positive constants C_1, α, and C_2 such that,

$$\int_{\Theta_{k-1,n,\varepsilon}} dG_{k-1,n}(\theta) \ge 1 - C_1 \exp\left\{-\alpha_1 n^{\varepsilon/3} \log n\right\} \ge C_2.$$

Thus,

$$\int_{\Theta_{n,\varepsilon}} dG_{k-1,n}(\theta) \ge C_2. \qquad (4.37)$$

Substituting this last inequality back into (4.36) we write,

$$\widehat{W}^{\dagger\dagger(n)}_{T} \le \prod_{k=1}^{Tn} C_2^{-1} \times \prod_{k=1}^{Tn} \left(1 + \int_{\Theta_{n,\varepsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1}\theta dG_{k-1,n}(\theta)\right)$$

which after assignment $C_3 = \prod_{k=1}^{Tn} C_2^{-1}$ becomes,

$$\widehat{W}_T^{\dagger\dagger(n)} \leq C_3 \prod_{k=1}^{Tn} \left(1 + \int_{\Theta_{n,\varepsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1} \theta dG_{k-1,n}(\theta)\right).$$

Now split integration over regions $\Theta_{k-1,n,\varepsilon}$ and $\Theta_{k-1,n,\varepsilon}/\Theta_{n,\varepsilon}$. Hence we write,

$$\begin{aligned}\widehat{W}_T^{\dagger\dagger(n)} &\leq C_3 \prod_{k=1}^{Tn} \left(1 + \int_{\Theta_{k-1,n,\varepsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1} \theta dG_{k-1,n}(\theta)\right. \\ &\left. + \int_{\Theta_{k-1,n,\varepsilon}/\Theta_{n,\varepsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1} \theta dG_{k-1,n}(\theta)\right). \end{aligned} \tag{4.38}$$

Let us bound the integrand $\left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1}\theta$ on region $\Theta_{n,\varepsilon}$. Note that,

$$\begin{aligned}\left|\left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1}\theta\right| &\leq \left\|\widetilde{X}_{k/n} - \mathbf{1}\right\| \left\|A_{\lfloor (k-1)/n \rfloor + 1}\theta\right\| \\ &\leq \frac{\sqrt{m} L_P (1 + \log n)}{\sqrt{n}} L_{A,d}^{1/2} \|\theta\|. \end{aligned}$$

But since $\theta \in \Theta_{n,\varepsilon}$ it follows that $\|\theta\| \leq \left(\frac{n^{\varepsilon/3} \log n}{4\sqrt{m} L_{A,d}^{1/2} L_P}\right)$ and therefore,

$$\begin{aligned}\left|\left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1}\theta\right| &\leq \frac{n^{\varepsilon/3} \log n (1 + \log n)}{4\sqrt{n}} \\ &\leq \frac{3}{4} \text{ for sufficiently small } \varepsilon. \end{aligned} \tag{4.39}$$

Using this bound on the second integral of (4.38) we write,

$$\begin{aligned}\widehat{W}_T^{\dagger\dagger(n)} &\leq C_3 \prod_{k=1}^{Tn} \left(1 + \int_{\Theta_{k-1,n,\varepsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1} \theta dG_{k-1,n}(\theta)\right. \\ &\left. + \frac{3}{4} \int_{\Theta_{k-1,n,\varepsilon}/\Theta_{n,\varepsilon}} dG_{k-1,n}(\theta)\right). \\ &= C_3 \prod_{k=1}^{Tn} \left(1 + \int_{\Theta_{k-1,n,\varepsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1} \theta dG_{k-1,n}(\theta)\right) \\ &\times \left[1 + \left(\frac{3}{4} \int_{\Theta_{k-1,n,\varepsilon}/\Theta_{n,\varepsilon}} dG_{k-1,n}(\theta)\right)\right. \\ &\left. \times \left(1 + \int_{\Theta_{k-1,n,\varepsilon}} \left(\widetilde{X}_{k/n} - \mathbf{1}\right)' A_{\lfloor (k-1)/n \rfloor + 1} \theta dG_{k-1,n}(\theta)\right)^{-1}\right]. \end{aligned} \tag{4.40}$$

Now note that from (4.25) that,

$$\widehat{W}_T^{\dagger(n)} = \prod_{k=1}^{Tn}\left(1+\int_{\Theta_{k-1,n,\epsilon}}\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\theta dG_{k-1,n}(\theta)\right).$$

Substituting this into (4.40) we get,

$$\begin{aligned}\widehat{W}_T^{\dagger\dagger(n)} &\leq C_3\widehat{W}_T^{\dagger(n)}\prod_{k=1}^{Tn}\left[1+\left(\frac{3}{4}\int_{\Theta_{k-1,n,\epsilon}/\Theta_{n,\epsilon}} dG_{k-1,n}(\theta)\right)\right.\\ &\quad\left.\times\left(1+\int_{\Theta_{k-1,n,\epsilon}}\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\theta dG_{k-1,n}(\theta)\right)^{-1}\right]. \quad (4.41)\end{aligned}$$

Now take the last factor in this expression. Note that by virtue of (4.39),

$$\begin{aligned}&\left(1+\int_{\Theta_{k-1,n,\epsilon}}\left(\widetilde{X}_{k/n}-\mathbf{1}\right)' A_{\lfloor (k-1)/n\rfloor+1}\theta dG_{k-1,n}(\theta)\right)^{-1}\\ &\leq \left(1-\frac{3}{4}\int_{\Theta_{k-1,n,\epsilon}} dG_{k-1,n}(\theta)\right)^{-1}\\ &\leq \left(1-\frac{3}{4}\right)^{-1}=4.\end{aligned}$$

Substituting this into (4.41) we get,

$$\widehat{W}_T^{\dagger\dagger(n)} \leq C_3\widehat{W}_T^{\dagger(n)}\prod_{k=1}^{Tn}\left[1+3\int_{\Theta_{k-1,n,\epsilon}/\Theta_{n,\epsilon}} dG_{k-1,n}(\theta)\right].$$

Now note that,

$$\int_{\Theta_{k-1,n,\epsilon}/\Theta_{n,\epsilon}} dG_{k-1,n}(\theta) \leq \int_{\Theta_{k-1,n,\epsilon}/\mathbf{R}^d} dG_{k-1,n}(\theta).$$

But by virtue of (4.29) $\int_{\Theta_{k-1,n,\epsilon}/\mathbf{R}^d} dG_{k-1,n}(\theta)$ is in turn bounded above by,$C_1\exp\left\{-\alpha_1 n^{\epsilon/3}\log n\right\}$. Using this bound and absorbing the constant 3 into C_1 we write,

$$\begin{aligned}\widehat{W}_T^{\dagger\dagger(n)} &\leq C_3\widehat{W}_T^{\dagger(n)}\prod_{k=1}^{Tn}\left[1+C_1\exp\left\{-\alpha_1 n^{\epsilon/3}\log n\right\}\right]\\ &= C_3\widehat{W}_T^{\dagger(n)}\left[1+C_1\exp\left\{-\alpha_1 n^{\epsilon/3}\log n\right\}\right]^{Tn}\\ &\leq C_3\widehat{W}_T^{\dagger(n)}\exp\left\{C_1 Tn\exp\left\{-\alpha_1 n^{\epsilon/3}\log n\right\}\right\}.\end{aligned}$$

Upon dividing through by $\widehat{W}_T^{\dagger(n)}$, the lemma is proven. ■

The final lemma to be considered links the wealth of Procedure 4 to that of the continuously traded Procedure 1 of Section 3.4.1.

Lemma 4.2.4 *Suppose the Universality Conditions are satisfied. Let $\widehat{W}_T^{\dagger\dagger(n)}$ be the wealth of Procedure 3 after T time periods and let $\widehat{W}_T$.be the wealth of Procedure 1 after T time periods. Then for the ε of Procedure 3, there exists positive constants C and α depending on ε such that,*

$$\frac{\widehat{W}_T}{\widehat{W}_T^{\dagger\dagger(n)}} \leq C \exp\left\{T\epsilon_n\right\},$$

where,

$$\epsilon_n = \alpha \log^6 n / n^{1/2-\varepsilon}.$$

Proof. Start with the wealth $\widehat{W}_T^{\dagger\dagger(n)}$ written as,

$$\begin{aligned}
\widehat{W}_T^{\dagger\dagger(n)} &= W_0 \prod_{k=1}^{Tn} \widehat{b}_{(k-1)/n}^{\dagger\dagger(n)'} X_{k/n} \\
&= W_0 \prod_{k=1}^{Tn} \frac{\int_{\Theta_{n,\varepsilon}} b'\left(\theta, s_{\lfloor (k-1)/n \rfloor + 1}\right) X_{k/n} dG_{(k-1),n}\left(\theta\right)}{\int_{\Theta_{n,\varepsilon}} dG_{(k-1),n}\left(\theta\right)}. \qquad (4.42)
\end{aligned}$$

Here it is understood that,

$$b\left(\theta, s_{\lfloor (k-1)/n \rfloor+1}\right) = \left(b_0\left(\theta, s_{\lfloor (k-1)/n \rfloor+1}\right), A_{\lfloor (k-1)/n \rfloor+1}\theta\right)$$

where $b_0\left(\theta, s_{\lfloor (k-1)/n \rfloor+1}\right) = 1 - \mathbf{1}\cdot\left(A_{\lfloor (k-1)/n \rfloor+1}\theta\right)$ and $\mathbf{1}$ is a vector of m 1's. Now recall from (4.20) that,

$$\int_{\Theta_{n,\varepsilon}} dG_{(k-1),n}\left(\theta\right) \equiv \frac{\int_{\Theta_{n,\varepsilon}} E_{(k-1),n}\left(\theta\right) d\pi\left(\theta\right)}{\int_{\mathbf{R}^d} E_{k,n}\left(\theta\right) d\pi\left(\theta\right)}.$$

where $E_{k,n}\left(\theta\right)$ is given in (4.12). We would now like to apply Lemma 4.1.1 to allow us to write $E_{k,n}\left(\theta\right) = W_{k/n}^{(n)}(\theta) \exp\left\{\varepsilon_{k/n}\left(\theta\right)\right\}$ where $W_{k/n}^{(n)}(\theta)$ is the wealth achieved by discrete procedure $b\left(\theta, s_{\lfloor k/n \rfloor+1}\right)$and $\varepsilon_{k/n}\left(\theta\right)$ is the remainder term in Lemma 4.1.1's expression for $W_{k/n}^{(n)}(\theta)$. In order for the lemma to hold it is required that $\left\|A_{\lfloor k/n \rfloor+1}\theta\right\|_1 < B\left(n\right)$ for some bound $B\left(n\right)$ which in turn is bounded by $B\left(n\right) \leq \sqrt{n}/L_P\left(1+\log\left(n\right)\right)$. Note that for $\theta \in \Theta_{n,\varepsilon}$ that

$$\begin{aligned}
\left\|A_{\lfloor k/n \rfloor+1}\theta\right\|_1 &\leq L_{A,d}^{1/2}\left\|\theta\right\|_1 \\
&\leq \sqrt{m}L_{A,d}^{1/2}\left\|\theta\right\| \\
&\leq \sqrt{m}L_{A,d}^{1/2}\frac{n^{\varepsilon/3}\log n}{4\sqrt{m}L_{A,d}^{1/2}L_P} \\
&= \frac{n^{\varepsilon/3}\log n}{4L_P} \qquad (4.43) \\
&< \frac{\sqrt{n}}{L_P\left(1+\log\left(n\right)\right)} \text{ for sufficiently small } \varepsilon.
\end{aligned}$$

Thus upon setting $B(n) = \frac{n^{\epsilon/3}\log n}{L_P}$ it follows that the lemma holds and we can write $E_{k,n}(\theta) = W_{k/n}^{(n)}(\theta)\exp\left\{\varepsilon_{k/n}(\theta)\right\}$. Returning to (4.42) we write,

$$
\begin{aligned}
&\widehat{W}_T^{\dagger\dagger(n)} \\
= &\; W_0 \prod_{k=1}^{Tn} \frac{\int_{\Theta_{n,\epsilon}} b'\left(\theta, s_{\lfloor (k-1)/n \rfloor + 1}\right) X_{k/n} dG_{(k-1),n}(\theta)}{\int_{\Theta_{n,\epsilon}} dG_{(k-1),n}(\theta)} \\
= &\; W_0 \prod_{k=1}^{Tn} \frac{\int_{\Theta_{n,\epsilon}} b'\left(\theta, s_{\lfloor (k-1)/n \rfloor + 1}\right) X_{k/n} E_{(k-1),n}(\theta)\, d\pi(\theta)}{\int_{\Theta_{n,\epsilon}} E_{(k-1),n}(\theta)\, d\pi(\theta)} \\
= &\; W_0 \prod_{k=1}^{Tn} \frac{\int_{\Theta_{n,\epsilon}} b'\left(\theta, s_{\lfloor (k-1)/n \rfloor + 1}\right) X_{k/n} W_{(k-1)/n}^{(n)}(\theta)\exp\left\{\varepsilon_{(k-1)/n}(\theta)\right\} d\pi(\theta)}{\int_{\Theta_{n,\epsilon}} W_{(k-1)/n}^{(n)}(\theta)\exp\left\{\varepsilon_{(k-1)/n}(\theta)\right\} d\pi(\theta)} \\
= &\; W_0 \prod_{k=1}^{Tn} \frac{\int_{\Theta_{n,\epsilon}} W_{k/n}^{(n)}(\theta)\exp\left\{\varepsilon_{(k-1)/n}(\theta)\right\} d\pi(\theta)}{\int_{\Theta_{n,\epsilon}} W_{(k-1)/n}^{(n)}(\theta)\exp\left\{\varepsilon_{(k-1)/n}(\theta)\right\} d\pi(\theta)} \\
= &\; W_0 \prod_{k=1}^{Tn} \frac{\int_{\Theta_{n,\epsilon}} W_{k/n}^{(n)}(\theta)\exp\left\{\varepsilon_{k/n}(\theta)\right\}\exp\left\{\varepsilon_{(k-1)/n}(\theta) - \varepsilon_{k/n}(\theta)\right\} d\pi(\theta)}{\int_{\Theta_{n,\epsilon}} W_{(k-1)/n}^{(n)}(\theta)\exp\left\{\varepsilon_{(k-1)/n}(\theta)\right\} d\pi(\theta)}. \quad (4.44)
\end{aligned}
$$

Now recall from Lemma 4.1.1 that $\left|\varepsilon_{k/n}(\theta) - \varepsilon_{(k-1)/n}(\theta)\right|$ is $O\left(B(n)^3(1+\log n)^3/n^{3/2}\right)$. Thus there exists a constant C such that $\left|\varepsilon_{k/n}(\theta) - \varepsilon_{(k-1)/n}(\theta)\right| \le C\,B(n)^3(1+\log n)^3/n^{3/2}$. Since we declared previously that $B(n) = \frac{n^{\epsilon/3}\log n}{L_P}$ this implies (after constant $1/L_P$ absorbed into C) that, $\left|\varepsilon_{k/n}(\theta) - \varepsilon_{(k-1)/n}(\theta)\right| \le C\,\log^3 n\,(1+\log n)^3/n^{3/2-\varepsilon}$. Increasing C slightly we can write,

$$\left|\varepsilon_{k/n}(\theta) - \varepsilon_{(k-1)/n}(\theta)\right| \le C\log^6 n/n^{3/2-\varepsilon}.$$

Thus returning to (4.44) we write,

$$
\begin{aligned}
\widehat{W}_T^{\dagger\dagger(n)} &\ge W_0 \prod_{k=1}^{Tn} \exp\left\{-C\log^6 n/n^{3/2-\varepsilon}\right\} \frac{\int_{\Theta_{n,\epsilon}} W_{k/n}^{(n)}(\theta)\exp\left\{\varepsilon_{k/n}(\theta)\right\} d\pi(\theta)}{\int_{\Theta_{n,\epsilon}} W_{(k-1)/n}^{(n)}(\theta)\exp\left\{\varepsilon_{(k-1)/n}(\theta)\right\} d\pi(\theta)} \\
&\ge W_0 \exp\left\{-CT\log^6 n/n^{1/2-\varepsilon}\right\} \prod_{k=1}^{Tn} \frac{\int_{\Theta_{n,\epsilon}} W_{k/n}^{(n)}(\theta)\exp\left\{\varepsilon_{k/n}(\theta)\right\} d\pi(\theta)}{\int_{\Theta_{n,\epsilon}} W_{(k-1)/n}^{(n)}(\theta)\exp\left\{\varepsilon_{(k-1)/n}(\theta)\right\} d\pi(\theta)} \\
&= W_0 \exp\left\{-CT\log^6 n/n^{1/2-\varepsilon}\right\} \frac{\int_{\Theta_{n,\epsilon}} W_T^{(n)}(\theta)\exp\left\{\varepsilon_T(\theta)\right\} d\pi(\theta)}{\int_{\Theta_{n,\epsilon}} W_0\, d\pi(\theta)}.
\end{aligned}
$$

Now recall that from Lemma 4.1.1 that $\varepsilon_{k/n}(\theta)$ is $O\left(B^3(n)(1+\log n)^3 k/n^{3/2}\right)$. Moreover because we are only considering $k \le Tn$ we can find a constant C' such that $\left|\varepsilon_{k/n}(\theta)\right| \le C'B^3(n)(1+\log n)^3 T/n^{1/2}$. Also because $B(n) = \frac{n^{\epsilon/3}\log n}{L_P}$ it follows (after constant $1/L_P$

is absorbed into C') that $\left|\varepsilon_{k/n}(\theta)\right| \leq C'T\log^3 n\,(1+\log n)^3/n^{1/2-\varepsilon}$. Again by increasing C' slightly it follows that $\left|\varepsilon_{k/n}(\theta)\right| \leq C'T\log^6 n/n^{1/2-\varepsilon}$. Thus we write,

$$\begin{aligned}\widehat{W}_T^{\dagger\dagger(n)} &\geq W_0 \exp\left\{-CT\log^6 n/n^{1/2-\varepsilon}\right\} \times \\ &\exp\left\{-C'T\log^6 n/n^{1/2-\varepsilon}\right\} \frac{\int_{\Theta_{n,\varepsilon}} W_T^{(n)}(\theta)\,d\pi(\theta)}{\int_{\Theta_{n,\varepsilon}} W_0 d\pi(\theta)}.\end{aligned}$$

However after reassigning C to be $C+C'$ we write,

$$\begin{aligned}\widehat{W}_T^{\dagger\dagger(n)} &\geq W_0 \exp\left\{-CT\log^6 n/n^{1/2-\varepsilon}\right\} \frac{\int_{\Theta_{n,\varepsilon}} W_T^{(n)}(\theta)\,d\pi(\theta)}{W_0\int_{\Theta_{n,\varepsilon}} d\pi(\theta)} \\ &= \exp\left\{-CT\log^6 n/n^{1/2-\varepsilon}\right\} \frac{\int_{\Theta_{n,\varepsilon}} W_T^{(n)}(\theta)\,d\pi(\theta)}{\pi(\Theta_{n,\varepsilon})}.\end{aligned}$$

Since $\Theta_{n,\varepsilon}$ grows with n it follows that there is some positive constant C_2^{-1} such that $\pi(\Theta_{n,\varepsilon}) > C_2^{-1}$. Hence,

$$\widehat{W}_T^{\dagger\dagger(n)} \geq C_2 \exp\left\{-CT\log^6 n/n^{1/2-\varepsilon}\right\} \int_{\Theta_{n,\varepsilon}} W_T^{(n)}(\theta)\,d\pi(\theta). \tag{4.45}$$

Now we need to compare the wealth $W_T^{(n)}(\theta)$ achieved by rebalancing n times a period with the wealth $W_T(\theta)$ achieved by rebalancing continuously. Examining the ratio $W_T(\theta)/W_T^{(n)}(\theta)$ we write,

$$\begin{aligned}\frac{W_T(\theta)}{W_T^{(n)}(\theta)} &= \exp\left\{\sum_{\tau=1}^{T}\mu_\tau^{\dagger\prime}A_{\lceil\tau\rceil}\theta + \frac{1}{2}\sum_{\tau=1}^{T}\sum_{j=1}^{m}K_{\tau,j,j}^{\dagger}\left(A_{\lceil\tau\rceil}\theta\right)_j \right.\\ &\left. -\sum_{\tau=1}^{T}\frac{1}{2}\theta A_{\lceil\tau\rceil}' K_\tau^{\dagger} A_{\lceil\tau\rceil}\theta\right\} \times \\ &\exp\left\{-\sum_{\tau=1}^{T}\mu_\tau^{\dagger\prime}A_{\lceil\tau\rceil}\theta - \frac{1}{2}\sum_{\tau=1}^{T}\sum_{j=1}^{m}K_{\tau,j,j}^{\dagger(n)}\left(A_{\lceil\tau\rceil}\theta\right)_j \right.\\ &\left. +\sum_{\tau=1}^{T}\frac{1}{2}\theta A_{\lceil\tau\rceil}' K_\tau^{\dagger(n)} A_{\lceil\tau\rceil}\theta + \varepsilon_T(\theta)\right\} \\ &= \exp\left\{\frac{1}{2}\sum_{\tau=1}^{T}\sum_{j=1}^{m}\left(K_{\tau,j,j}^{\dagger} - K_{\tau,j,j}^{\dagger(n)}\right)\left(A_{\lceil\tau\rceil}\theta\right)_j \right.\\ &\left. -\sum_{\tau=1}^{T}\frac{1}{2}\theta A_{\lceil\tau\rceil}'\left(K_\tau^{\dagger} - K_\tau^{\dagger(n)}\right)A_{\lceil\tau\rceil}\theta + \varepsilon_T(\theta)\right\}.\end{aligned} \tag{4.46}$$

Here $(A_{\lceil\tau\rceil}\theta)_j$ denotes the jth entry of $A_{\lceil\tau\rceil}\theta$. Now we bound the absolute value of the exponent. Note that,

$$\begin{aligned}
&\left|\frac{1}{2}\sum_{\tau=1}^{T}\sum_{j=1}^{m}\left(K^{\dagger}_{\tau,j,j}-K^{\dagger(n)}_{\tau,j,j}\right)(A_{\lceil\tau\rceil}\theta)_j-\sum_{\tau=1}^{T}\frac{1}{2}\theta A'_{\lceil\tau\rceil}\left(K^{\dagger}_{\tau}-K^{\dagger(n)}_{\tau}\right)A_{\lceil\tau\rceil}\theta+\varepsilon_{k/n}(\theta)\right| \\
\leq\ & \frac{1}{2}\sum_{\tau=1}^{T}\left\|\text{diag}\left(K^{\dagger}_{\tau,j,j}-K^{\dagger(n)}_{\tau,j,j}\right)\right\|\left\|A_{\lceil\tau\rceil}\theta\right\|+\frac{1}{2}\sum_{\tau=1}^{T}\left\|\text{diag}\left(K^{\dagger}_{\tau,j,j}-K^{\dagger(n)}_{\tau,j,j}\right)\right\|\left\|A_{\lceil\tau\rceil}\theta\right\|^2 \\
\leq\ & \frac{1}{2}L_{A,d}^{1/2}\|\theta\|\sum_{\tau=1}^{T}|\lambda|_{\max}\left(K^{\dagger}_{\tau,j,j}-K^{\dagger(n)}_{\tau,j,j}\right)+\frac{1}{2}L_{A,d}\|\theta\|^2\sum_{\tau=1}^{T}|\lambda|_{\max}\left(K^{\dagger}_{\tau,j,j}-K^{\dagger(n)}_{\tau,j,j}\right) \\
\leq\ & \frac{1}{2}L_{A,d}^{1/2}\frac{L'_k}{\sqrt{n}}T\|\theta\|+\frac{1}{2}L_{A,d}\frac{L'_k}{\sqrt{n}}T\|\theta\|^2.
\end{aligned}$$

But recall that $\|\theta\|\leq\left(\frac{n^{\varepsilon/3}\log n}{4\sqrt{m}L_{A,d}^{1/2}L_P}\right)$ for $\theta\in\Theta_{n,\varepsilon}$, so for some constant C_3 the above is bounded by,

$$\leq C_3T\log^2 n/n^{1/2-2\varepsilon/3},$$

which in turn after readjustment of C_3 is upper bounded by,

$$\leq C_3T\log^6 n/n^{1/2-\varepsilon}.$$

Using this in conjunction with (4.46) we write,

$$\frac{W_T(\theta)}{W_T^{(n)}(\theta)}\leq\exp\left\{C_3T\log^6 n/n^{1/2-\varepsilon}\right\}\quad\text{for }\theta\in\Theta_{n,\varepsilon}.$$

Returning to (4.45) we write,

$$\begin{aligned}
\widehat{W}_T^{\dagger\dagger(n)} &\geq C_2\exp\left\{-CT\log^6 n/n^{1/2-\varepsilon}\right\}\int_{\Theta_{n,\varepsilon}}W_T^{(n)}(\theta)\,d\pi(\theta) \\
&\geq C_2\exp\left\{-CT\log^6 n/n^{1/2-\varepsilon}\right\}\exp\left\{-C_3T\log^6 n/n^{1/2-\varepsilon}\right\}\times \\
&\quad\int_{\Theta_{n,\varepsilon}}W_T^{(n)}(\theta)\frac{W_T(\theta)}{W_T^{(n)}(\theta)}d\pi(\theta).
\end{aligned}$$

Resetting $C=C+C_3$ we get,

$$\widehat{W}_T^{\dagger\dagger(n)}\geq C_2\exp\left\{-CT\log^6 n/n^{1/2-\varepsilon}\right\}\int_{\Theta_{n,\varepsilon}}W_T(\theta)\,d\pi(\theta). \tag{4.47}$$

The last step in the proof changes the domain of integration from $\Theta_{n,\varepsilon}$ to $\mathbf{R}^d$. Note that,

$$\int_{\Theta_{n,\varepsilon}}W_T(\theta)\,d\pi(\theta)=\left(\int_{\Theta_{n,\varepsilon}}\frac{W_T(\theta)\,d\pi(\theta)}{\int_{\mathbf{R}^d}W_T(\theta)\,d\pi(\theta)}\right)\int_{\mathbf{R}^d}W_T(\theta)\,d\pi(\theta).$$

Now recall that

$$dG_T = \frac{W_T(\theta)\, d\pi(\theta)}{\int_{\mathbf{R}^d} W_T(\theta)\, d\pi(\theta)}$$

is a Gaussian measure where

$$G_T \sim N\left(\Psi_T, \left(\Omega_T^{-1} + \Lambda^{-1}\right)^{-1}\right).$$

We have already argued in the proofs of Lemma 4.2.2 and Lemma 4.2.3 that for the related measure,

$$G_{k,n} \sim N\left(\Psi_{k/n}, \left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1}\right)$$

there exists a constants C_4, C and α_1 independent of k and n such that,

$$\int_{\Theta_{n,\epsilon}} dG_{k,n}(\theta) \geq 1 - C\exp\left\{-\alpha_1 n^{\epsilon/3}\log n\right\} \geq C_4.$$

I claim that these same results hold if $G_{k,n}$ is replaced by G_T. It is just a matter of reworking the old proofs. Nothing substantial changes. Again we can find a constant C_4 independent of T such that,

$$\int_{\Theta_{n,\epsilon}} \frac{W_T(\theta)\, d\pi(\theta)}{\int_{\mathbf{R}^d} W_T(\theta)\, d\pi(\theta)} \geq \int_{\Theta_{n,\epsilon}} dG_T(\theta) \geq C_4.$$

This in turn implies that

$$\int_{\Theta_{n,\epsilon}} W_T(\theta)\, d\pi(\theta) \geq C_4 \int_{\mathbf{R}^d} W_T(\theta)\, d\pi(\theta) .. \tag{4.48}$$

Substituting (4.48) back into (4.47) we write,

$$\widehat{W}_T^{\dagger\dagger(n)} \geq C_2 C_4 \exp\left\{-CT\log^6 n/n^{1/2-\epsilon}\right\} \int_{\Theta_{n,\epsilon}} W_T(\theta)\, d\pi(\theta).$$

Upon resetting $C_2 = C_2 C_4$ and noting that,

$$\widehat{W}_T = \int_{\mathbf{R}^d} W_T(\theta)\, d\pi(\theta),$$

we conclude that,

$$\widehat{W}_T^{\dagger\dagger(n)} \geq C_2 \exp\left\{-CT\log^6 n/n^{1/2-\epsilon}\right\} \widehat{W}_T.$$

Thus the lemma is proven. ■

4.2.3 Main Theorem

Now we have the all the lemmas in place to prove the major theorem for this chapter. The theorem gives a bound on how close the wealth of Procedure 2 comes to that of the target wealth of the continuously traded target class $\left(\mathbf{R}^d, S, A(s)\theta\right)$.

Theorem 4.2.5 *Assume that the Universality Conditions hold. Let $\widehat{W}_T^{(n)}$ be the wealth achieved as of time T by Procedure 2. Let W_T^* be the wealth achieved by the best in hindsight strategy in the continuously traded target class $\left(\mathbf{R}^d, S, A(s)\theta\right)$. Then for $n > \exp\left\{24\sqrt{m}L_{A,d}^{1/2}L_P L_\Psi\right\}$ and any $\varepsilon > 0$ we can find constants α and C such that,*

$$\frac{W_T^*}{\widehat{W}_T^{(n)}} \leq CT^d \exp\left\{\alpha T/n^{1/2-\varepsilon}\right\}.$$

Proof. Note that,

$$\frac{W_T^*}{\widehat{W}_T^{(n)}} = \frac{W_T^*}{\widehat{W}_T} \times \frac{\widehat{W}_T}{\widehat{W}_T^{\dagger\dagger(n)}} \times \frac{\widehat{W}_T^{\dagger\dagger(n)}}{\widehat{W}_T^{\dagger(n)}} \times \frac{\widehat{W}_T^{\dagger(n)}}{\widehat{W}_T^{(n)}}.$$

Using Theorem 3.3.2 we know that $\frac{W_T^*}{\widehat{W}_T} = O\left(T^d\right)$ so,

$$\frac{W_T^*}{\widehat{W}_T^{(n)}} = O\left(T^d\right) \times \frac{\widehat{W}_T}{\widehat{W}_T^{\dagger\dagger(n)}} \times \frac{\widehat{W}_T^{\dagger\dagger(n)}}{\widehat{W}_T^{\dagger(n)}} \times \frac{\widehat{W}_T^{\dagger(n)}}{\widehat{W}_T^{(n)}}.$$

Setting the ε parameter of Procedure 4 to 2ε we use Lemma 4.2.4 to note that

$$\begin{aligned}\frac{\widehat{W}_T}{\widehat{W}_T^{\dagger\dagger(n)}} &\leq C\exp\left\{\alpha T \log^6 n/n^{1/2-2\varepsilon}\right\} \\ &\leq C\exp\left\{\alpha T/n^{1/2-\varepsilon}\right\} \text{ after some increase in } \alpha.\end{aligned}$$

Hence,

$$\frac{W_T^*}{\widehat{W}_T^{(n)}} \leq O\left(T^d\right)\exp\left\{\alpha T/n^{1/2-\varepsilon}\right\} \times \frac{\widehat{W}_T^{\dagger\dagger(n)}}{\widehat{W}_T^{\dagger(n)}} \times \frac{\widehat{W}_T^{\dagger(n)}}{\widehat{W}_T^{(n)}}. \tag{4.49}$$

Again setting the ε parameter of Procedure 3 to 2ε we use Lemma 4.2.3 to note that,

$$\frac{\widehat{W}_T^{\dagger\dagger(n)}}{\widehat{W}_T^{\dagger(n)}} \leq C'\exp\left\{CTn\exp\left\{-\alpha n^{2\varepsilon/3}\log n\right\}\right\}.$$

But $\exp\left\{-\alpha n^{2\varepsilon/3}\log n\right\}$ decreases faster than any polynomial power and therefore the above is further bounded (after appropriate readjustment of C) by,

$$\frac{\widehat{W}_T^{\dagger\dagger(n)}}{\widehat{W}_T^{\dagger(n)}} \leq C' \exp\left\{CT/n^{1/2-\varepsilon}\right\}.$$

This bound is absorbed into (4.49) and after appropriate increase in α we write,

$$\frac{W_T^*}{\widehat{W}_T^{(n)}} \leq O\left(T^d\right)\exp\left\{\alpha T/n^{1/2-\varepsilon}\right\} \times \frac{\widehat{W}_T^{\dagger(n)}}{\widehat{W}_T^{(n)}}. \tag{4.50}$$

Finally note from Lemma 4.2.2 that

$$\frac{\widehat{W}_T^{\dagger(n)}}{\widehat{W}_T^{(n)}} \leq \exp\left\{CTn\exp\left\{-\alpha n^{\varepsilon/3}\log n\right\}\right\}. \tag{4.51}$$

Just as before we realize that for an appropriate increase of the α in (4.50) we can take the bound (4.51) and absorb it into bound (4.50). Hence,

$$\frac{W_T^*}{\widehat{W}_T^{(n)}} \leq O\left(T^d\right)\exp\left\{\alpha T/n^{1/2-\varepsilon}\right\}.$$

The theorem follows immediately. ■

Theorem 4.2.5 shows us how we might use Procedure 2 to be universal with respect to target wealth W_T^*. The first observation we make is that for fixed n,

$$\begin{aligned}\lim_{T\to\infty}\sup_{P_t}\frac{1}{T}\log\frac{W_T^*}{\widehat{W}_T^{(n(T))}} &\leq \lim_{T\to\infty}\frac{1}{T}\log\left(CT^d\exp\left\{\alpha T/n^{1/2-\varepsilon}\right\}\right)\\ &= \lim_{T\to\infty}\alpha/n^{1/2-\varepsilon}.\end{aligned} \tag{4.52}$$

Here $\sup_{P_t}$ represents the supremum over all price paths satisfying the minimal path and universality conditions. From (4.52) we see that by using Procedure 2 we can get arbitrarily close to the optimal growth rate by choosing n to be sufficiently large. To be actually universal though, we need to trade Procedure 2 at increasingly smaller intervals. By making $n(T)$ an increasing function of T such that $\lim_{T\to\infty} n(T) = \infty$, we see that,

$$\lim_{T\to\infty}\sup_{P_t}\frac{1}{T}\log\frac{W_T^*}{\widehat{W}_T^{(n(T))}} \leq \lim_{T\to\infty}\alpha/n(T)^{1/2-\varepsilon} = 0,$$

which implies that Procedure 2 is now universal. The second observation we make is that if $n(T) \geq T^{2+\delta}$ for some $\delta > 0$ then the wealth ratio $W_T^*/\widehat{W}_T^{(n)}$ becomes $O\left(T^d\right)$. To see this note that,

$$\begin{aligned}\frac{W_T^*}{\widehat{W}_T^{(n(T))}} &\leq CT^d\exp\left\{\alpha T/n^{1/2-\varepsilon}\right\}\\ &\leq CT^d\exp\left\{\alpha T/T^{1-2\varepsilon+\delta/2-\varepsilon\delta}\right\}.\end{aligned}$$

By choosing $\varepsilon < \frac{\delta}{2(2+\delta)}$, it follows that $1 - 2\varepsilon + \delta/2 - \varepsilon\delta > 1$ and hence $T/T^{1-2\varepsilon+\delta/2-\varepsilon\delta}$ is upper bounded by some constant implying that,

$$\frac{W_T^*}{\widehat{W}_T^{(n(T))}} \leq CT^d.$$

Hence the wealth of the procedure would be within a polynomial bound of the target wealth if $n(T) \geq T^{2+\delta}$.

4.3 Experiments with NYSE Data

4.3.1 The Procedures

In this section we examine the behavior of Procedure 2 on actual stock data. Recall the definition of Procedure 2,

Procedure 2 *At time k/n invest according to $\widehat{b}_{k/n}^{(n)} = \left(\widehat{b}_{k/n,0}^{(n)}, \widehat{\widetilde{b}}_{k/n}^{(n)}\right)$ where,*

$$\widehat{\widetilde{b}}_{k/n}^{(n)} = A_{\lfloor k/n \rfloor + 1} \Psi_{k/n},$$

and $\widehat{b}_{k/n,0}^{(n)} = 1 - \sum_{j=1}^{m} \widehat{b}_{k/n,j}^{(n)}$. Here $\Psi_{k/n}$ is computed through the relations,

$$\begin{aligned}
\Psi_{k/n} &= \left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1} \left(\sum_{\tau \in I(k/n)} u_\tau + \Lambda^{-1}\lambda\right), \\
\Omega_{k/n}^{-1} &\equiv A_1' K_1^{\dagger(n)} A_1 + \cdots + A_{\lceil k/n \rceil - 1}' K_{\lceil k/n \rceil - 1}^{\dagger(n)} A_{\lceil k/n \rceil - 1} + A_{\lceil k/n \rceil}' K_{k/n}^{\dagger(n)} A_{\lceil k/n \rceil}, \\
u_\tau &\equiv A_{\lceil \tau \rceil}' \left[\mu_\tau^\dagger + \frac{1}{2} \text{diag} K_\tau^{\dagger(n)}\right],
\end{aligned}$$

where Λ is an arbitrary positive definite $d \times d$ matrix and λ is an arbitrary vector in $\mathbf{R}^d$.

We have applied the procedure to two different target classes. The first target class corresponds to the class of constant rebalanced portfolio. The second target class is similar to the example given in section 3.2.2 where we split wealth between a constant rebalanced portfolio and another portfolio based on past price observations. We discuss these classes in more detail below.

Constant Rebalanced Portfolio Recall that a constant rebalanced portfolio invests in the same portfolio each period. Thus portfolio strategies in this class are enumerated by selecting $\theta \in \mathbf{R}^m$ and investing in,

$$b(\theta) = \left(1 - \sum_{j=1}^{m} \theta_j, \theta_1, \dots, \theta_m\right)'$$

at the start of each period. We also note that the stock component of $b(\theta)$ is writable as,

$$\widetilde{b}(\theta) = \mathbf{I}\,\theta,$$

where $\mathbf{I}$ is the $m \times m$ identity matrix. For the purposes of computing Procedure 2 we would therefore compute the stock component of Procedure 2 as,

$$\widetilde{\widetilde{b}}^{(n)}_{k/n} = \mathbf{I}\,\Psi_{k/n} = \Psi_{k/n}.$$

Here $\Psi_{k/n}$ would be computed as described in the definition of Procedure 2.

Growth Optimal/CRP Split For this class of strategies we consider splitting wealth between a constant rebalanced portfolio (CRP) and another portfolio based on an estimated covariance matrix κ_τ of the log prices of the m stocks over a 250 day moving window and log price drift δ_τ computed on a 15 day moving window. The form of this latter portfolio is given by $s_\tau = (s_{\tau,0}, \widetilde{s}_\tau)$, where

$$\widetilde{s}_\tau = \kappa_{\tau-1}^{-1}\left(\delta_{\tau-1} + \frac{1}{2}\mathrm{diag}\left(\kappa_{\tau-1}\right)\right),$$

represents the stock component of the portfolio. This portfolio is essentially an estimate of the growth optimal portfolio of Corollary 3.1.2. Assuming that the covariances of log prices are relatively stable over long periods and that the drift of the last 15 days is a reasonable proxy for the drift of the next few days, this portfolio should be a reasonable estimate of the growth optimal portfolio of the coming period.

We will assume that a new iteration of $\widetilde{s}_\tau$ is computed each week. Thus the number of periods T is equal to the number of weeks in our investment horizon. For $\theta \in \mathbf{R}^{m+1}$ members of this target class set the stock component,

$$\widetilde{b}(\theta, s_\tau) = \theta_{m+1} s'_\tau + (\theta_1, \dots, \theta_m)$$

at the start of each week. During the course of the week we treat this portfolio as a constant rebalanced portfolio and rebalance at the start of each day for a total of 5 rebalances per week (i.e. $n = 5$).

Stock	Ending Wealth Relative
WFC	6.92
XON	3.33
BCC	1.06

Table 4.1: Wealth relatives achieved over the 9 year period.

In order to track the wealth of the best strategy in hindsight among members of this class we are motivated to use Procedure 2. In order to facilitate the procedure's calculation we note that the stock component of the class's portfolio mapping is equal to,

$$\tilde{b}(\theta, s_\tau) = A_\tau \theta$$

where A_τ is the $(m+1) \times d$ matrix given by,

$$A_\tau = (s_\tau, \mathbf{I}) \text{ where } \mathbf{I} \text{ is the } m \times m \text{ identity matrix.}$$

Given A_τ, we see that the stock component of Procedure 2 is immediately computable through the formula,

$$\tilde{\tilde{b}}^{(n)}_{k/n} = A_{\lfloor k/n \rfloor + 1} \Psi_{k/n}.$$

4.3.2 The Data

In order to gauge how Procedure 2 performs in an applied setting we implemented it with respect to the two target classes described above using 9 years of stock data from the NYSE. The data consists of daily closing prices from January 17th, 1989 to January 17th, 1998 for the 3 stocks of Wells Fargo (WFC), Boise Cascade (BCC), and Exxon (XON). The data has been adjusted in the usual manner for dividends and stock splits. Figure 4.1 shows how the wealth relatives of each stock evolved over time. We can see from the graph that Wells Fargo outperformed the other two stocks for most of the nine year period. To be exact, Table 4.1 shows that one dollar placed originally in Wells Fargo became \$6.92 by January 17th, 1998. In contrast the performance of Boise Cascade was rather flat earning just 6 cents on the dollar. Exxon was somewhere in between with one dollar becoming \$3.33.

Procedure 2 was implemented on this data with respect to the two target classes described in the previous section. In both cases the implementation used a "prior" mean of $\lambda = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ so as to initially spread wealth evenly among the three stocks. For the CRP class we set $\Lambda = \mathbf{I}$ and for the Growth Optimal/CRP Split class $\Lambda = 1000 \times \mathbf{I}$. Rebalances were executed daily and side information was updated once a week (i.e. $n = 5$ rebalances a

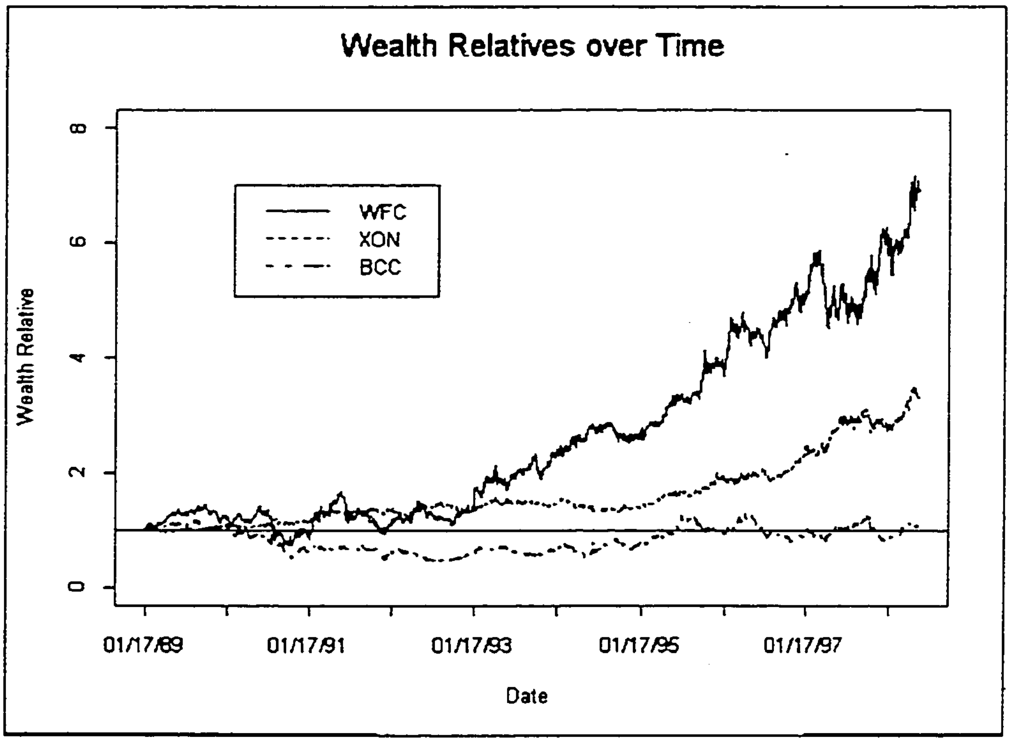


Figure 4.1: Wealth relatives of the constituent stocks over the 9 year period.

week for approximately $T = 470$ weeks). In order to implement Procedure 2 with respect to the Growth Optimal/CRP Split class it was necessary to employ some data prior to January 17th. 1989. This was necessary to compute the initial values of κ and δ. All procedures started trading on January 17th, 1989.

As a means of comparison, a short-selling version of Cover's universal portfolio was also implemented. Specifically we estimated the performance of.

$$\tilde{\hat{b}}_i = \frac{\int_{\mathbf{R}^3} b W_{i-1}(b) d\mu(b)}{\int_{\mathbf{R}^3} W_{i-1}(b) d\mu(b)}, \quad \text{with } \mu = N\left(\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \mathbf{I}\right). \tag{4.53}$$

This portfolio is very similar to the proposed constant rebalanced portfolio implementation of Procedure 2. In particular we expect from the analysis conducted in the previous two chapters that both procedures converge to the same continuous time procedure as rebalances become more frequent. Hence performances should be similar. Note however that CRP implementation of Procedure 2 has advantages over (4.53) when it comes to computation. Whereas Procedure 2 is computed through a m^2 algorithm (m being the number of stocks) procedure (4.53) is calculated via analytic or numeric techniques having computational order growing exponentially in m. While calculation of (4.53) is feasible in our example of three stocks and cash the number of calculations can quickly become unweildy for larger groups of stock.

The evolution of wealth relatives of Procedure 2 applied to the constant rebalanced portfolio class as well as that for Procedure 2 applied to the Growth Optimal/CRP split class and that for our adaptation of Cover's universal procedure are presented in Figure 4.2. While there are supposedly three portfolios represented in the figure we can only see two lines. This is because the CRP implementation of Procedure 2 and the short selling Cover type procedure implement nearly identically portfolios at each step. Differences are so small that the wealth relatives at any point are nearly identical. Over the considered time period the maximal difference between the two wealth relatives is never more than 0.025.

This confirms our earlier claim that the two procedures converge as rebalancing become more frequent. Clearly with this data there is little difference in the performance of the two procedures under daily rebalancing. We reiterate however that there is a big difference in order of computation. In this instance Procedure 2 is computable in m^2 time versus the exponential in m time of (4.53) so we are likely to prefer the former for practical purposes.

This similarity of performance is also seen in Table 4.2 where we see both procedures obtain the same final wealth relative. We also note from Table 4.2 and Figure 4.2 that the two constant rebalance portfolio procedures beat the Growth Optimal/CRP procedure. The use of extra side information didn't seem to help us much here. Even so. we should

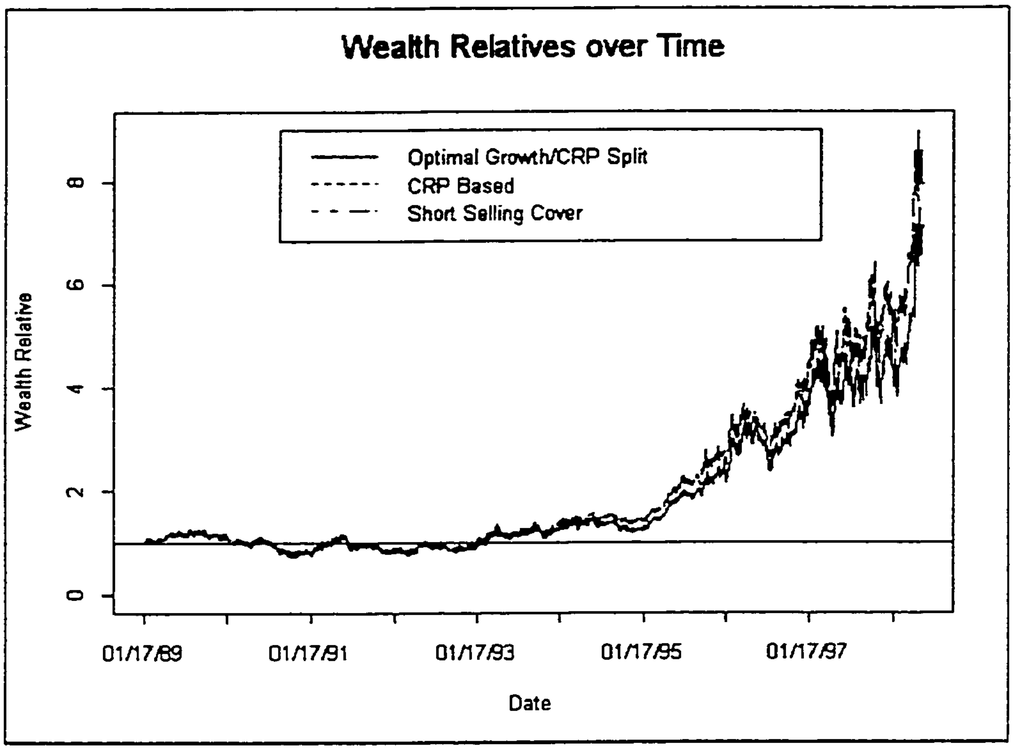


Figure 4.2: Wealth relatives of the three universal procedures over time.

Procedure	Ending Wealth Relative
CRP Based	7.95
Growth Optimal/CRP Split	7.15
Short Selling Cover	7.95

Table 4.2: Wealth relatives achieved over the 9 year period.

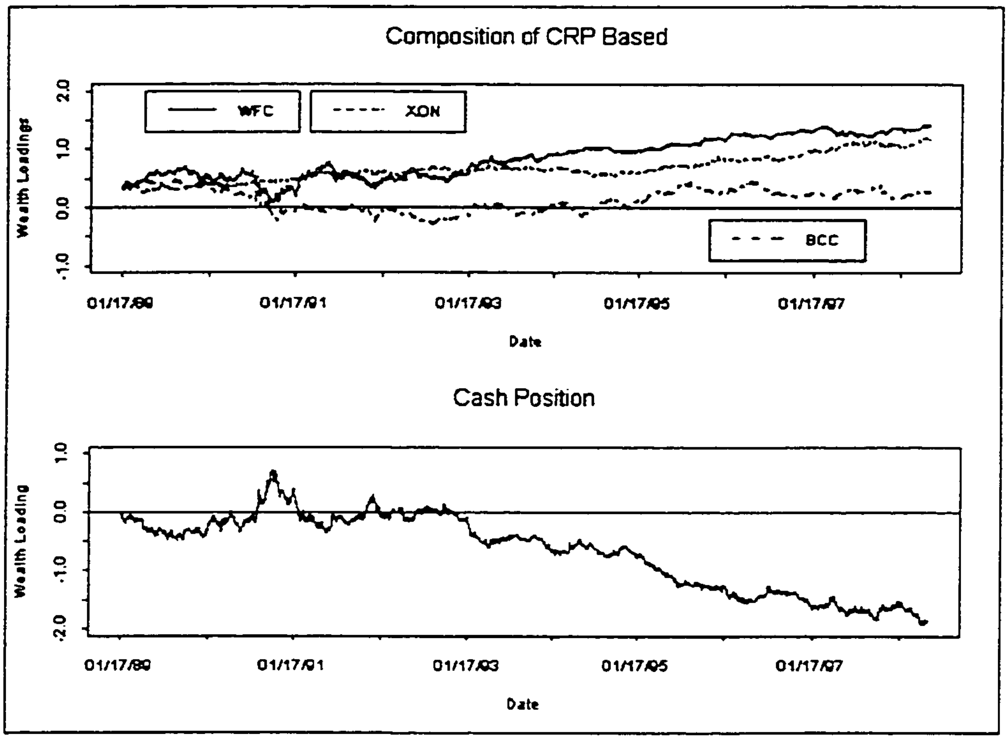


Figure 4.3: Positions in Stock and Cash of Procedure 3 wrt to CRP target class.

note that all three strategies managed to outperformed the best stock (Wells Fargo) over the given period .

Figure 4.3 shows the evolution of the proportions of wealth in stock and cash used by the CRP based implementation of Procedure 2. To gain more understanding of what the graph is telling us, we note that at the beginning of 1989 each "loading" is equal to 1/3. This indicates that wealth has been spread evenly initially among the three stocks. Also note that the "wealth loading" for Wells Fargo (WFC) near the end of the time frame is slightly above one. This indicates that we have invested somewhat more than our entire net worth in Wells Fargo. In order to achieve such a position it is necessary to borrow money. Indeed the second graph of Figure 4.3 labelled cash position shows that we are borrowing about twice our net worth by the end of the period. Thus implementing the procedure requires a fair amount of leverage. We can also see that the procedure has shorted Boise Cascade at various points. This occurs when the BCC line falls below the 0 level.

Figure 4.4 exhibits the wealth loadings for Procedure 2 applied to the Growth Optimal/CRP Split class. We can see that the loadings are significantly more variable than the CRP based procedure. However the end positions and amount of leverage used are

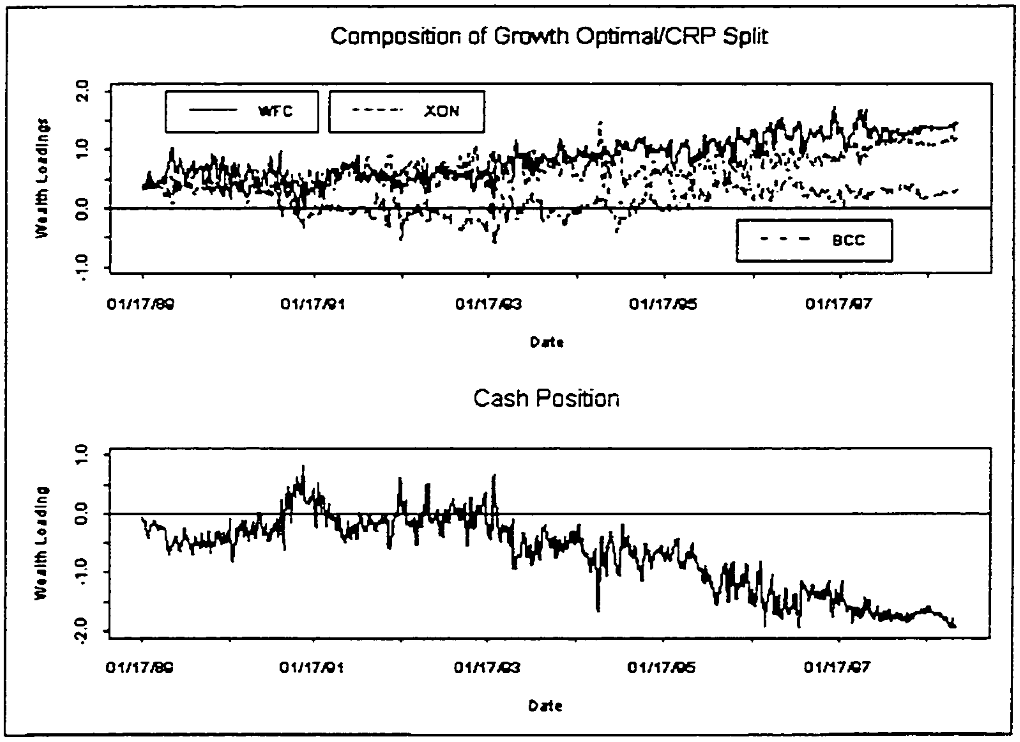


Figure 4.4: Wealth Loadings of Procedure 3 applied to the Growth Optimal/CRP Split class.

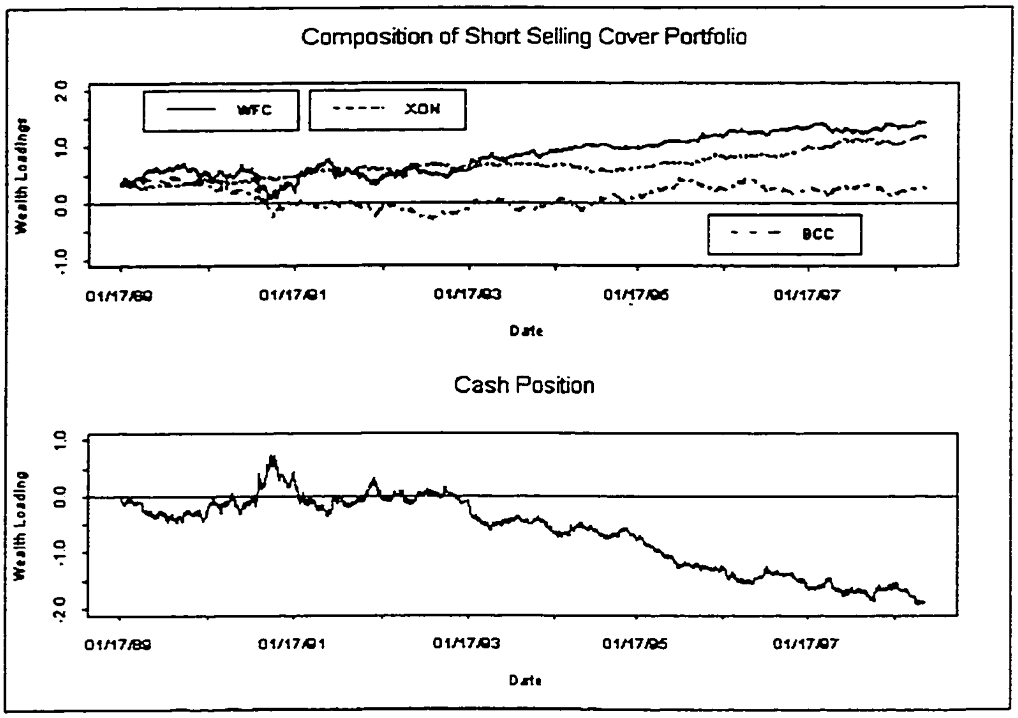


Figure 4.5: Loadings used by Cover's Universal Procedure.

roughly the same as before.

Figure 4.5 shows the loadings used by the short selling Cover portfolio. As expected the figure is a near replica of Figure 4.3. The loadings of the two procedures are nearly identical. Again we see that Wells Fargo is the most heavily weighted stock followed by Exxon and Boise Cascade.

4.4 A Note on Computation

As we have mentioned before, one of the main reasons for considering Procedure 2 is its ease of computation. Recall that the discrete procedures developed in Chapter 2 (of which Cover's universal procedure is a subcase) are computable in steps that grow geometrically with the dimensionality d of the parameter space. However in order to compute Procedure 2 we claim that only a constant factor of d^2 steps is needed at each iteration.

Recall that the Procedure 2 is computed through the formula

$$\tilde{\tilde{b}}^{(n)}_{k/n} = A_{\lfloor k/n \rfloor + 1} \Psi_{k/n},$$

where

$$\begin{aligned}
\Psi_{k/n} &= \left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1}\left(\sum_{\tau\in I(k/n)} u_\tau + \Lambda^{-1}\lambda\right), \\
\Omega_{k/n}^{-1} &\equiv A_1' K_1^{\dagger(n)} A_1 + \cdots + A'_{\lceil k/n\rceil-1} K^{\dagger(n)}_{\lceil k/n\rceil-1} A_{\lceil k/n\rceil-1} + A'_{\lceil k/n\rceil} K^{\dagger(n)}_{k/n} A_{\lceil k/n\rceil}, \\
u_\tau &\equiv A'_{\lceil\tau\rceil}\left[\mu_\tau^\dagger + \frac{1}{2}\mathrm{diag}K_\tau^{\dagger(n)}\right].
\end{aligned}$$

Assume that the dimensionality of the parameter space d is greater than the number of stocks m. We can see that at each step the computation of $\Psi_{k/n}$ depends on the computed value of $\left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1}$. Suppose that $\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}$ has been stored. We claim that,

$$\begin{aligned}
&\left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1} \\
= &\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1} \\
&-\frac{\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1} A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}}{1+\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}}. \qquad (4.54)
\end{aligned}$$

where $\delta_{k/n}$ is the m-dimensional incremental log-drift vector,

$$\delta_{k/n} = Z_{k/n} - Z_{(k-1)/n}.$$

To see that (4.54) holds, first note that,

$$\left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right) = \left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right) + A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}. \qquad (4.55)$$

Thus if (4.54) multiplied by (4.55) is the identity matrix then we know that (4.54) is an identity for $\left(\Omega_{k/n}^{-1} + \Lambda^{-1}\right)^{-1}$. Multiplying (4.54) by (4.55) we write,

$$\begin{aligned}
&I - \frac{A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}}{1+\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}} + A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1} \\
&-\frac{A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}}{1+\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}} \\
= &I + \frac{A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}}{1+\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}} \\
&-\frac{A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}}{1+\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega_{(k-1)/n}^{-1} + \Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}}.
\end{aligned}$$

But $\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega^{-1}_{(k-1)/n}+\Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}$ is a scalar so the above equals,

$$\begin{aligned} &= I+\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega^{-1}_{(k-1)/n}+\Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}\left(\frac{A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega^{-1}_{(k-1)/n}+\Lambda^{-1}\right)^{-1}}{1+\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega^{-1}_{(k-1)/n}+\Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}}\right. \\ &\quad\left.-\frac{A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega^{-1}_{(k-1)/n}+\Lambda^{-1}\right)^{-1}}{1+\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega^{-1}_{(k-1)/n}+\Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}}\right) \\ &= I. \end{aligned}$$

So (4.54) holds. As for the order of computation of (4.54) note that the vector $A'_{\lceil k/n\rceil}\delta_{k/n}$ takes on order dm steps to compute. The matrix $A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}$ is computed in d^2 steps. The numerator of the second term of (4.54),

$$\left(\Omega^{-1}_{(k-1)/n}+\Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega^{-1}_{(k-1)/n}+\Lambda^{-1}\right)^{-1}$$

is also computed in d^2 steps as is the denominator,

$$1+\delta'_{k/n}A_{\lceil k/n\rceil}\left(\Omega^{-1}_{(k-1)/n}+\Lambda^{-1}\right)^{-1}A'_{\lceil k/n\rceil}\delta_{k/n}.$$

Thus (4.54) and $\left(\Omega^{-1}_{k/n}+\Lambda^{-1}\right)^{-1}$ are computable within a factor of d^2 steps.

The only component remaining to be calculated is $\sum_{\tau\in I(k/n)}u_\tau$. If $\sum_{\tau\in I((k-1)/n)}u_\tau$ has been stored, note that we can compute $\sum_{\tau\in I(k/n)}u_\tau$ via the formula

$$\sum_{\tau\in I(k/n)}u_\tau=\sum_{\tau\in I((k-1)/n)}u_\tau+\eta_{k/n},$$

where

$$\eta_{k/n}=A'_{\lceil k/n\rceil}\left[\delta_{k/n}+\frac{1}{2}\text{diag}\delta_{k/n}\delta'_{k/n}\right]$$

and $\delta_{k/n}$ is the m-dimensional incremental log-drift vector,

$$\delta_{k/n}=Z_{k/n}-Z_{(k-1)/n}.$$

We see that $\eta_{k/n}$ is computable within a factor of dm steps and hence $\sum_{\tau\in I(k/n)}u_\tau$ is computable within a factor of dm steps. Finally we see that $\Psi_{k/n}$ is the product of the $d\times d$ matrix $\left(\Omega^{-1}_{k/n}+\Lambda^{-1}\right)^{-1}$ and d dimensional vector $\left(\Omega^{-1}_{k/n}v_{k/n}+\Lambda^{-1}\lambda\right)$. Such a matrix multiplication is accomplished in d^2 steps. In all, the most steps required by any step of $\Psi_{k/n}$'s calculation is on the order d^2. Thus in general, $\Psi_{k/n}$ is computable within a factor of d^2 steps.

Chapter 5

Appendix

5.1 Proof of Theorem 3.1.1

Proof. Set an arbitrary time horizon T. We work directly with the definition of W_t as the limit of wealths $W_t^{(n)}(b)$ achieved by rebalancing portfolio b a total of n times over the time horizon T. Recall from equation (3.1) that,

$$W_t(b) = \lim_{n\to\infty} W_t^{(n)}(b) = \lim_{n\to\infty} W_0 \prod_{k=1}^{\lfloor nt/T \rfloor} \left(1 + \sum_{j=1}^{m} b_j \left(\frac{P_{kT/n,j}}{P_{(k-1)T/n,j}} - 1\right)\right).$$

To simplify exposition it will be useful to define "returns",

$$R_{k,j}(n) \equiv \left(\frac{P_{kT/n,j}}{P_{(k-1)T/n,j}} - 1\right).$$

In terms of the $R_{k,j}(n)$ we can write $W_t^{(n)}(b)$ as,

$$\begin{aligned} W_t(b) &= \lim_{n\to\infty} W_0 \prod_{k=1}^{\lfloor nt/T \rfloor} \left(1 + \sum_{j=1}^{m} b_j R_{k,j}(n)\right) \\ &= \lim_{n\to\infty} W_0 \exp\left\{\sum_{k=1}^{\lfloor nt/T \rfloor} \log\left(1 + \sum_{j=1}^{m} b_j R_{k,j}(n)\right)\right\}. \end{aligned}$$

Consider the expansion $\log(1+x) = x - x^2/2 + x^3/3(1+c)^3$ (for some c between x and 0). Using this expansion along with the substitution $x = \sum_{j=1}^{m} b_j R_{k,j}(n)$ we write,

$$\begin{aligned} W_t(b) &= W_0 \exp\left\{ \lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \sum_{j=1}^{m} b_j R_{k,j}(n) - \frac{1}{2} \lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \left(\sum_{j=1}^{m} b_j R_{k,j}(n) \right)^2 \right. \\ &\left. + \lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \frac{1}{3(1+c_k)^3} \left(\sum_{j=1}^{m} b_j R_{k,j}(n) \right)^3 \right\}. \end{aligned} \tag{5.1}$$

Here, c_k is some number between 0 and $\sum_{j=1}^{m} b_j R_{k,j}(n)$. We work on bounding each of the limits in the exponent of (5.1). First note that from minimal path assumptions,

$$\begin{aligned} \left| \sum_{j=1}^{m} b_j R_{k,j}(n) \right| &\le \left\| \tilde{b} \right\|_1 \max_j |R_{k,j}(n)| \\ &\le \left\| \tilde{b} \right\|_1 \frac{L_P(1+\log n)}{\sqrt{n}}. \end{aligned}$$

Now use this bound on the third limit of (5.1). Note that,

$$\begin{aligned} & \lim_{n\to\infty} \left| \sum_{k=1}^{\lfloor nt/T \rfloor} \frac{1}{3(1+c_k)^3} \left(\sum_{j=1}^{m} b_j R_{k,j}(n) \right)^3 \right| \\ \le & \lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \left| \frac{1}{3(1+c_k)^3} \right| \left(\left\| \tilde{b} \right\|_1 \frac{L_P(1+\log n)}{\sqrt{n}} \right)^3. \end{aligned} \tag{5.2}$$

But note that $3(1+c_k)^3 \ge 3\left(1 - \left\| \tilde{b} \right\|_1 \frac{L_P(1+\log n)}{\sqrt{n}}\right)^3$ which is bounded below by some positive constant C for sufficiently large n. Thus we further upper bound (5.2) by,

$$\begin{aligned} &\le \lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \frac{1}{C} \left(\left\| \tilde{b} \right\|_1 \frac{L_P(1+\log n)}{\sqrt{n}} \right)^3 \\ &\le \lim_{n\to\infty} \frac{nt}{TC} \left\| \tilde{b} \right\|_1^3 L_P^3 (1+\log n)^3 n^{-3/2} \\ &= 0. \end{aligned}$$

Thus the third limit of (5.1) vanishes. Now consider the second limit of (5.1),

$$\lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \left(\sum_{j=1}^{m} b_j R_{k,j}(n) \right)^2.$$

Use the expansion $x = \log(1+x) + \frac{x^2}{2(1+c)^2}$ with x replaced by $R_{k,j}(n)$ to write,

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \left(\sum_{j=1}^{m} b_j R_{k,j}(n) \right)^2 \\
= &\lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \left(\sum_{j=1}^{m} b_j \log(1+R_{k,j}(n)) + \sum_{j=1}^{m} b_j \frac{R_{k,j}^2(n)}{2(1+c_{k,j})} \right)^2 \\
= &\lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \left(\sum_{j=1}^{m} b_j \log(1+R_{k,j}(n)) \right)^2 \\
&+ \lim_{n\to\infty} 2 \sum_{k=1}^{\lfloor nt/T \rfloor} \left(\sum_{j=1}^{m} b_j \log(1+R_{k,j}(n)) \sum_{j=1}^{m} b_j \frac{R_{k,j}^2(n)}{2(1+c_{k,j})} \right) \\
&+ \lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \left(\sum_{j=1}^{m} b_j \frac{R_{k,j}^2(n)}{2(1+c_{k,j})} \right)^2, \qquad (5.3)
\end{aligned}$$

where $c_{k,j}$ is some number between 0 and $R_{k,j}(n)$. We examine each limit of (5.3) in turn. First note that,

$$\log(1+R_{k,j}(n)) = \log\left(\frac{P_{kT/n,j}}{P_{(k-1)T/n,j}}\right) = Z_{kT/n,j} - Z_{(k-1)T/n,j}.$$

Thus we can write the first limit of (5.3) as,

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \left(\sum_{j=1}^{m} b_j \log(1+R_{k,j}(n)) \right)^2 \\
= &\lim_{n\to\infty} \sum_{k=1}^{\lfloor nt/T \rfloor} \sum_{i=1}^{m} \sum_{j=1}^{m} b_i b_j \left(Z_{kT/n,i} - Z_{(k-1)T/n,i}\right)\left(Z_{kT/n,j} - Z_{(k-1)T/n,j}\right) \\
= &\lim_{n\to\infty} \sum_{i=1}^{m} \sum_{j=1}^{m} b_i b_j \sum_{k=1}^{\lfloor nt/T \rfloor} \left(Z_{kT/n,i} - Z_{(k-1)T/n,i}\right)\left(Z_{kT/n,j} - Z_{(k-1)T/n,j}\right) \\
= &\lim_{n\to\infty} \sum_{i=1}^{m} \sum_{j=1}^{m} b_i b_j K_{t,i,j}^{(n)} \\
= &\lim_{n\to\infty} \widetilde{b}' K_t^{(n)} \widetilde{b} \\
= &\ \widetilde{b}' K_t \widetilde{b}.
\end{aligned}$$

As for the second limit of (5.3) note that,

$$\begin{aligned}
&\lim_{n\to\infty}\left|\sum_{k=1}^{\lfloor nt/T\rfloor}\left(\sum_{j=1}^{m} b_j \log\left(1+R_{k,j}(n)\right)\sum_{j=1}^{m} b_j \frac{R_{k,j}^2(n)}{2\left(1+c_{k,j}\right)}\right)\right| \\
\le &\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\left(\left|\sum_{j=1}^{m} b_j \log\left(1+R_{k,j}(n)\right)\right|\left|\sum_{j=1}^{m} b_j \frac{R_{k,j}^2(n)}{2\left(1+c_{k,j}\right)}\right|\right).
\end{aligned}$$

Now observe that,

$$\left|\log\left(1+R_{k,j}(n)\right)\right| \le \left|R_{k,j}^2(n)\right| \le L_P\left(1+\log n\right)/\sqrt{n}.$$

Also, since $c_{k,j}$ is between 0 and $R_{k,j}(n)$ it follows that $2\left(1+c_{k,j}\right) > 2\left(1-L_P\left(1+\log n\right)/\sqrt{n}\right)$. But for sufficiently large n this again lower bounded by some positive constant C. Thus we continue by writing,

$$\begin{aligned}
\le &\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\left(\left\|\tilde{b}\right\|_1 \frac{L_P\left(1+\log n\right)}{\sqrt{n}} \times \left\|\tilde{b}\right\|_1 \frac{L_P^2\left(1+\log n\right)^2}{Cn}\right) \\
\le &\lim_{n\to\infty}\frac{nt}{T}\left\|\tilde{b}\right\|_1^2 \frac{L_P^3\left(1+\log n\right)^3}{Cn^{3/2}} \\
= &\ 0.
\end{aligned}$$

Thus the second limit of (5.3) vanishes. As for the third limit, note that,

$$\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\left(\sum_{j=1}^{m} b_j \frac{R_{k,j}^2(n)}{2\left(1+c_{k,j}\right)}\right)^2 \le \lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\left(\left\|\tilde{b}\right\|_1 \frac{L_P^2\left(1+\log n\right)^2}{Cn}\right)^2.$$

Again C is a positive constant that lower bounds $2\left(1+c_{k,j}\right)$ for sufficiently large n. Continuing we write,

$$\begin{aligned}
\le &\lim_{n\to\infty}\frac{nt}{T}\left(\left\|\tilde{b}\right\|_1 \frac{L_P^2\left(1+\log n\right)^2}{Cn}\right)^2 \\
= &\ 0.
\end{aligned}$$

Thus the third limit of (5.3) also vanishes.

The only task remaining is to evaluate the first limit of (5.1). Using the expansion,

$$x = \log\left(1+x\right) + x^2/2 - x^3/\left(1+c\right)^3$$

(for some c between 0 and x) and substituting $R_{k,j}(n)$ for x we write the first limit of (5.1) as,

$$\begin{aligned}
&\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m} b_j R_{k,j}(n) \\
= &\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m} b_j \log\left(1+R_{k,j}(n)\right) + \frac{1}{2}\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m} b_j R_{k,j}^2(n) \\
&-\frac{1}{3}\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m} b_j \frac{R_{k,j}^3(n)}{(1+c_{k,j})^3}.
\end{aligned} \tag{5.4}$$

Now we evaluate each limit of (5.4). Start with the third limit. Note that for sufficiently large n there is some positive constant C lower bounding $(1+c_{k,j})^3$. Thus we write,

$$\begin{aligned}
\lim_{n\to\infty}\left|\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m} b_j \frac{R_{k,j}^3(n)}{(1+c_{k,j})^3}\right| &\le \lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\left\|\widetilde{b}\right\|_1 \frac{L_P^3(1+\log n)^3}{Cn^{3/2}} \\
&\le \lim_{n\to\infty}\frac{nt}{T}\left\|\widetilde{b}\right\|_1 \frac{L_P^3(1+\log n)^3}{Cn^{3/2}} \\
&= 0.
\end{aligned}$$

Hence the third limit of (5.4) vanishes. As for the first limit of (5.4) note that,

$$\begin{aligned}
\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m} b_j \log\left(1+R_{k,j}(n)\right) &= \lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m} b_j\left(Z_{kT/n,i}-Z_{(k-1)T/n,i}\right) \\
&= \lim_{n\to\infty}\sum_{j=1}^{m} b_j \sum_{k=1}^{\lfloor nt/T\rfloor}\left(Z_{kT/n,i}-Z_{(k-1)T/n,i}\right) \\
&= \lim_{n\to\infty}\sum_{j=1}^{m} b_j\left(Z_{\lfloor nt/T\rfloor\times T/n,i}-Z_{0,i}\right) \\
&= \lim_{n\to\infty}\widetilde{b}'\mu_{\lfloor nt/T\rfloor\times T/n} \\
&= \widetilde{b}'\mu_t.
\end{aligned}$$

The only limit remaining to be evaluated is the second limit of (5.4). Again we must break up this limit into parts and evaluate each part. Again use the expansion $x=\log(1+x)+\frac{x^2}{2(1+c)^2}$

with x replaced by $R_{k,j}(n)$ to write,

$$\begin{aligned}\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m}b_jR_{k,j}^2(n) &= \lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m}b_j\left(\log(1+R_{k,j}(n))+\frac{R_{k,j}^2(n)}{2(1+c_{k,j})^2}\right)^2\\ &= \lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m}b_j\log^2(1+R_{k,j}(n))\\ &\quad+\lim_{n\to\infty}2\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m}b_j\log(1+R_{k,j}(n))\frac{R_{k,j}^2(n)}{2(1+c_{k,j})^2}\\ &\quad+\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m}b_j\left(\frac{R_{k,j}^2(n)}{2(1+c_{k,j})^2}\right)^2.\end{aligned}\tag{5.5}$$

The first limit of (5.5) evaluates to,

$$\begin{aligned}\lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m}b_j\log^2(1+R_{k,j}(n)) &= \lim_{n\to\infty}\sum_{j=1}^{m}b_j\sum_{k=1}^{\lfloor nt/T\rfloor}\left(Z_{kT/n,i}-Z_{(k-1)T/n,i}\right)^2\\ &= \lim_{n\to\infty}\sum_{j=1}^{m}b_jK_{t,j,j}^{(n)}\\ &= \sum_{j=1}^{m}K_{t,j,j}b_j.\end{aligned}$$

As for the second term, note that,

$$\begin{aligned}&\lim_{n\to\infty}\left|2\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m}b_j\log(1+R_{k,j}(n))\frac{R_{k,j}^2(n)}{2(1+c_{k,j})^2}\right|\\ \le\ &\lim_{n\to\infty}2\sum_{k=1}^{\lfloor nt/T\rfloor}\left\|\tilde{b}\right\|_1\frac{L_P(1+\log n)}{\sqrt{n}}\frac{L_P^2(1+\log n)^2}{Cn}.\end{aligned}$$

Here C is some positive constant that lower bounds $2(1+c_{k,j})^2$ for sufficiently large n. Continuing we write,

$$\begin{aligned}&\le \lim_{n\to\infty}2\frac{nt}{T}\left\|\tilde{b}\right\|_1\frac{L_P^3(1+\log n)^3}{Cn^{3/2}}\\ &= 0.\end{aligned}$$

Thus the second limit of (5.5) vanishes. Finally for the third limit of (5.5) we write,

$$\begin{aligned}\lim_{n\to\infty}\left|\sum_{k=1}^{\lfloor nt/T\rfloor}\sum_{j=1}^{m} b_j\left(\frac{R_{k,j}^2(n)}{2\left(1+c_{k,j}\right)^2}\right)^2\right| &\le \lim_{n\to\infty}\sum_{k=1}^{\lfloor nt/T\rfloor}\left\|\tilde{b}\right\|_1\left(\frac{L_P^2\left(1+\log n\right)^2}{Cn}\right)^2 \\ &\le \frac{nt}{T}\left\|\tilde{b}\right\|_1\left(\frac{L_P^2\left(1+\log n\right)^2}{Cn}\right)^2 \\ &= 0.\end{aligned}$$

Thus this limit vanishes too. The theorem follows immediately upon substituting the evaluations of all these limits back into (5.1). ■

5.2 Lemma Used in Theorems 2.3.2 and 3.3.2

Lemma 5.2.1 *Under the assumptions of Theorem 2.3.2, there exists a constant* $R > 0$ *independent of* n *such that* $\pi(\Theta_n^*) \geq Rn^{-d}$ *and* $\pi(\Theta^*_{\sqrt{n}}) \geq Rn^{-d/2}$. *Similarly, under the assumptions of Theorem 3.3.2 there exists a constant* $R > 0$ *independent of* n *such* $\pi(\Theta_t^*) \geq Rt^{-d}$ *and* $\pi(\Theta^*_{\sqrt{t}}) \geq R\left\lceil\sqrt{t}\right\rceil^{-d}$.

Proof. We start with the proof that $\pi(\Theta_n^*) \geq Rn^{-d}$. First note that because the Radon-Nikodym derivative of π is uniformly bounded above 0 by some $\delta > 0$ it follows that $\pi(\Theta_n^*) \geq \delta\text{Vol}(\Theta_n^*)$ where $\text{Vol}(\cdot)$ denotes the Lebesgue measure (or volume) of a set.

To bound $\text{Vol}(\Theta_n^*)$ it is useful to use the identity $\Theta_n^* = \Theta \bigcap B(\theta_n^*, \frac{1}{n})$ where, $B(\theta_n^*, \frac{1}{n}) = \{\theta \in \mathbf{R}^d : \|\theta - \theta_n^*\| \leq \frac{1}{n}\}$ is an appropriately centered closed ball in $\mathbf{R}^d$. The next step is to show that there exists a constant $C > 0$ independent of n for which $\text{Vol}(\Theta_n^*) \geq C\text{Vol}(B(\theta_n^*, \frac{\epsilon}{n}))$. In order to show this, we take a d-dimensional closed ball in Θ and consider the convex extension of the ball to θ_n^*. By examining the volume of the intersection between this extension and set Θ_n^* we are able to show that the necessary constant factor C exists.

To justify the existence of a d-dimensional closed ball in Θ we note that there exists d points in Θ such that there is no $d-1$ dimensional hyperplane containing all d points. If this were not the case, Θ would lie in a $d-1$ subspace and thus would have null Lebesgue measure which contradicts our assumptions. Thus, we could take the convex hull of these d non-planer points and by the convexity of Θ, the hull would be a subset of Θ. Clearly, a closed ball exists in such a hull and therefore also in Θ.

Suppose such a closed ball has center θ_0 and radius $r > 0$. Label it $B(\theta_0, r) \equiv \{\theta \in R^d \mid \|\theta - \theta_0\| \leq r\}$. Now define a convex extension to this ball. For parameter $\theta_n^* \in \Theta$ we define $C(\theta_n^*) = \{\theta : \theta = \lambda\theta_n^* + (1-\lambda)\theta',\ \lambda \in [0,1],\ \theta' \in B(\theta_0, r)\}$. By its definition

$C(\theta_n^*)$ is the convex hull of $B(\theta_0, r)$ and maximal parameter θ_n^*. The set can be visualized as an "ice cream cone" with tip θ_n^* and "scoop" $B(\theta_0, r)$. Since θ^* is contained in Θ as is $B(\theta_0, r)$, it follows from convexity of Θ that $C(\theta_n^*)$ is also in Θ.

Now consider the volume of the intersection between ball $B(\theta_n^*, 1/n)$ and cone $C(\theta_n^*)$. The cone $C(\theta_n^*)$ has been purposely defined to have its tip coincide with the center of $B(\theta_n^*, 1/n)$. For sufficiently large n, the radius of $B(\theta_n^*, 1/n)$ will be smaller than that of $B(\theta_0, r)$, (i.e. the ball atop $C(\theta_n^*)$). In this case, a geometric argument shows that,

$$\mathrm{Vol}\left(B(\theta_n^*, 1/n) \bigcap C(\theta_n^*)\right) \geq A(\theta_n^*) \mathrm{Vol}\left(B(\theta_n^*, 1/n)\right),$$

where $A(\theta_n^*)$ is the proportion of the surface area of $B(\theta_n^*, 1/n)$ contained in $C(\theta_n^*)$. As θ_n^* gets farther away from θ_0 (i.e. the center of the "scoop"), the cone narrows and $A(\theta_n^*)$ gets smaller. However it only vanishes completely when this distance between θ_n^* and θ_0 is infinite. Since Θ is compact, the distance is bounded and hence $A(\theta_n^*)$ is uniformly bounded above 0 for all $\theta_n^* \in \Theta$. Thus we can select constant $C > 0$ such that,

$$\mathrm{Vol}\left(B(\theta_n^*, 1/n) \bigcap C(\theta_n^*)\right) \geq C\mathrm{Vol}\left(B(\theta_n^*, 1/n)\right), \quad \forall \theta_n^* \in \Theta^*. \tag{5.6}$$

Although this inequality is justified only for n larger than some $N > 0$, we can make (5.6) hold for all n by defining C to be the lesser of,

$$\inf_{n \leq N} \left\{ \frac{\mathrm{Vol}\left(B(\theta_n^*, 1/n) \bigcap C(\theta_n^*)\right)}{\mathrm{Vol}\left(C(\theta^*)\right)} \right\},$$

and,

$$\inf_{\theta_n^* \in \Theta, n > N} A(\theta_n^*).$$

Both these infimums are strictly positive so we have $C > 0$ as required.

To end the proof we note that,

$$\begin{aligned}
\pi(\theta_{\epsilon/n}^*) &\geq \delta \mathrm{Vol}(\Theta_n^*) \\
&= \delta \mathrm{Vol}\left(B(\theta_n^*, 1/n) \bigcap \Theta\right) \\
&\geq \delta \mathrm{Vol}\left(B(\theta_n^*, 1/n) \bigcap C(\theta_n^*)\right) \\
&\geq \delta C \mathrm{Vol}\left(B(\theta_n^*, 1/n)\right) \\
&= \delta C K n^{-d} \text{ for some } K > 0 \\
&= R n^{-d} \text{ with } R = \delta C K.
\end{aligned}$$

The $n^{-d/2}$ result follows from an identical argument where n is replaced by $\sqrt{n}$ in the appropriate places. The results pertaining to Theorem 3.3.2 are proven immediately from the above arguments with virtually no modification. ■

Bibliography

[1] P. H. Algoet. Universal schemes for prediction, gambling, and portfolio selection. *Annals of Probability*, 20(2):901–941, 1992.

[2] P. H. Algoet and T. M. Cover. Asymptotic optimality and asymptotic equipartition properties of log-optimal investment. *Annals of Probability*, 16:876–898, 1988.

[3] R. Bell and T. M. Cover. Competitive optimality of logarithmic investment. *Mathematics of Operations Research*, 5(2):161–166, May 1980.

[4] L. Breiman. Optimal gambling systems for favorable games. *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pages 65–78, 1961.

[5] T. M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, January 1991.

[6] T. M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Transactions On Information Theory*, 42(2):348–363, March 1996.

[7] Thomas Cover and Joy Thomas. *Elements of Information Theory.* John Wiley and Sons, New York, 1991.

[8] Y. Singer D. P. Helmbold, R. E. Schapire and M. K. Warmuth. On-line portfolio selection using multiplicative updates. *Mathematical Finance*, 1(1):1–29, January 1998.

[9] Jr. J. L. Kelly. A new interpretation of the information rate. *Bell System Technical Journal*, 35:917–926, 1956.

[10] F. Jamshidian. Asymptotically optimal portfolios. *Mathematical Finance*, 2(2):131–150, 1992.

[11] D.C. Larson. *Growth Optimal Trading Strategies.* PhD thesis, Stanford University, Stanford, CA, 1986.

[12] H. Latane. Criteria for choice amongst risky ventures. *Journal of Political Economy*, 67:917–926, 1959.

[13] H. A. Latane. The geometric mean principle revisited. *Journal of Banking and Finance*, 2:395–398, 1978.

[14] H. A. Latane. The geometric mean criterion continued. *Journal of Banking and Finance*, 3:309–311, 1979.

[15] H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[16] R. C. Merton. Lifetime portfolio selection under uncertainty: The continuous time case. *Review of Economics and Statistics*, 51:247–257, 1969.

[17] R. C. Merton. Optimum consumption and portfolio rules in a continuous time model. *Journal of Economic Theory*, 3:373–413, 1971.

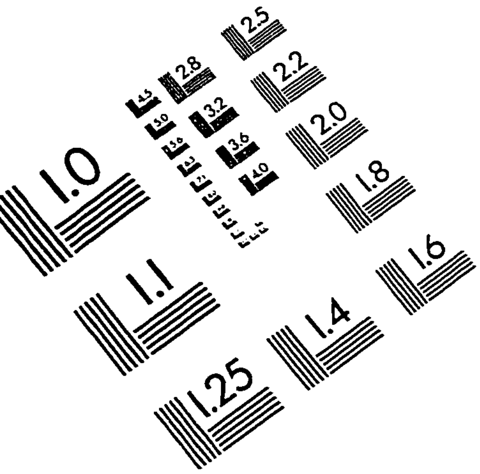
[18] T. Ophir. The geometric principle revisited. *Journal of Banking and Finance*, 2:103–107, 1978.

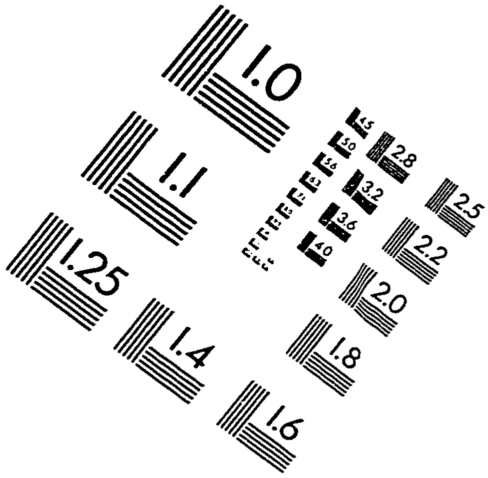
[19] T. Ophir. The geometric principle revisited: A reply to a reply. *Journal of Banking and Finance*, 3:305–308, 1979.

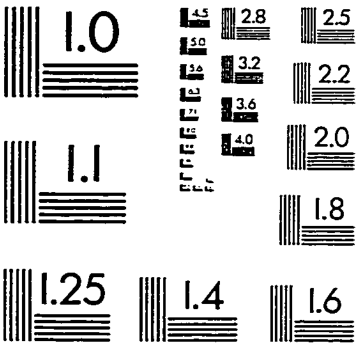
[20] P. A. Samuelson. Why we should not make mean log of wealth big though years to act are long. *Journal of Banking and Finance*, 3:305–307, 1979.

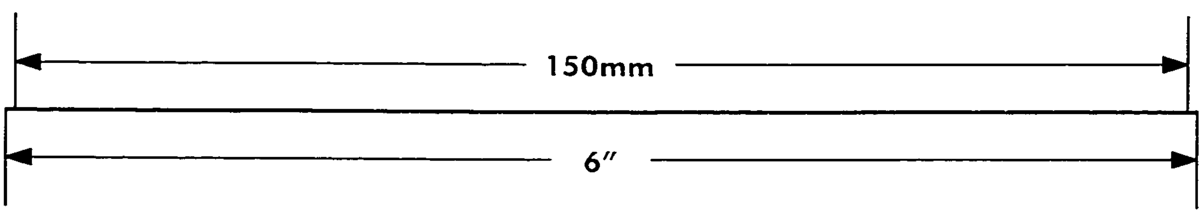
[21] W. F. Sharpe. *Investments*. Prentice-Hall, Englewood Cliffs, NJ, 1985.

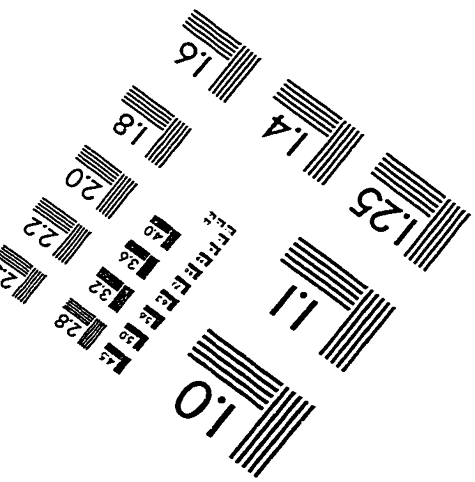
IMAGE EVALUATION TEST TARGET (QA-3)

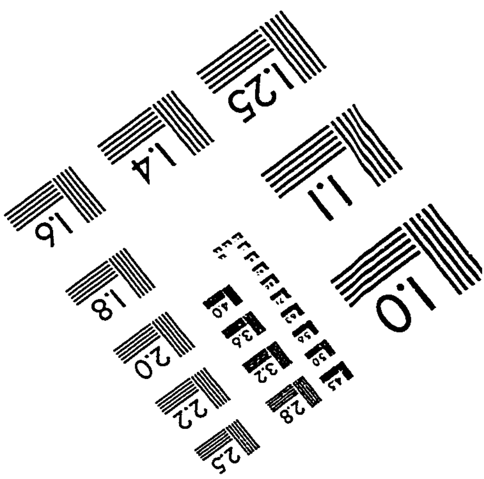












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989