University of Nevada, Reno

Dollar-Cost and Value Averaging
Theory and Applications

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematics.

by

Hayden T. Brown

Dr. Andrey Sarantsev, Thesis Advisor

May, 2021

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

HAYDEN BROWN

entitled

Dollar-Cost and Value Averaging
Theory and Applications

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Andrey Sarantsev, Ph.D.
Advisor

Ilya Zaliapin, Ph.D.
Committee Member

Aleksey Telyakovskiy, Ph.D.
Committee Member

Aleksey Kolpakov, Ph.D.
Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

May, 2021

Abstract

Dollar-cost and value averaging investing strategies are theoretically examined and applied for two asset classes: stocks and bonds. Theoretical expression of the cumulative distribution function, expectation and variance is developed for both strategies. Most of the obtained results are recursive, but some results for dollar-cost averaging are obtained in a closed form. In particular, closed form theoretical results are developed for dollar-cost averaging expectation, variance, Sharpe ratio and related optimization. Applications use annual data from the S&P Composite Index and 6-month bonds, providing results for expectation, variance, Sharpe ratio, related optimization and quantiles of returns. In applications, simulation is employed when theoretical results are unfeasible to compute. The results can be used by investors for selecting a desired risk-return balance. From a policy perspective, the results indicate that dollar-cost averaging is a viable investing strategy for large investment funds like university endowment funds.

Acknowledgements

I wish to thank my advisor, Andrey Sarantsev, for introducing me to the area of Financial Mathematics and Statistics. After his introduction, my interest grew quickly, and now I hope to have a career in this area. He is very encouraging, and such support makes for a great environment to do research in. I could not ask for a better advisor.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

1. INTRODUCTION

There are countless investing strategies available to the aspiring capitalist. The ideal strategy yields unusually high returns without exposing the investor to loss. Unfortunately, get-rich-quick strategies like this are hard to come by. In his famous book *A Random Walk Down Wall Street*, Burton Malkiel summarizes the large body of scientific evidence against the idea that frequent traders can consistently beat a diversified portfolio of broad-based index funds [18]. Following this line of thought, Malkiel advocates long-term investing strategies that focus on broad-based index funds. Furthermore, the proportion of investments in stocks, bonds and real estate funds can be adjusted to fit an investor's tolerance for risk. This paper selects two investing strategies that align with Malkiels advice and elaborates on the relationship between different asset-class proportions, risk and return.

The first investing strategy considered is dollar-cost averaging. In dollar-cost averaging, a constant amount is invested at set intervals. When investing in multiple asset classes, a proportion of that constant amount is invested in each asset class, at each time step. For example, consider investing \$100 at the beginning of each year: \$70 in stocks, \$20 in real estate and \$10 in bonds. Then repeat this for n years. This procedure is dollar-cost averaging with three asset classes. A similar procedure exists for dollar-cost averaging with any finite number of asset classes.

The second investing strategy considered is value averaging. In the context of this thesis, value averaging involves two asset classes. At the beginning of each investment period, a target amount is set for the riskier asset class to achieve at the end of the investment period. Upon reaching the end of the investment period, the target amount is achieved by buying or selling the riskier asset class. The less risky asset class is used as a reserve for when the target amount is surpassed. When the target amount is not surpassed, the less risky asset class is sold, with the profits being used to buy the riskier asset class and meet the target amount. If the target amount is still not met, a bottomless cash reserve is used to make the necessary additional purchases of the riskier asset class.

Here is an example of value averaging. Consider having \$100 invested in risky stocks and \$10 invested in riskless bonds at the beginning of the year. Set a target of \$106 for the end of the year. Suppose the stock investment rises to \$110. Then sell \$4 of stock and buy \$4 of bonds to end with \$106 in stocks, \$14 in bonds and \$110 total invested. Suppose the stock investment rises to \$102. Then sell \$4 of bonds and buy \$4 of stock to end with \$106 in stocks, \$6 in bonds and \$110 total invested. Suppose the stock investment drops to \$95. Then sell \$10 of bonds and buy \$11 of stock to end with \$106 in stocks, \$0 in bonds and \$111 total invested. Modify initial investments and the target amount to repeat for the next year.

1.1 Literature review

Dollar-cost averaging is a well-known investing strategy that has been practiced for over 70 years. In 1951, the Chicago Daily Tribune advocated for the dollar-cost averaging investing strategy [1]. The Tribune celebrated dollar-cost averaging because it makes no attempt at market timing, thereby helping investors avoid the pitfalls associated with trying to time the market. Under dollar-cost averaging with stock(s), investors buy more shares when market prices are low, and less shares with market prices are high. In effect, the average price payed per share is below average. As an example, it was shown that the dollar-cost averaging stragegy performed quite well with the single asset of EI du Pont de Nemours stock: an annual investment of $\$1,000$ for 26 years grew to $\$93,400$.

Morris (1959) provided a mathematical foundation for dollar-cost averaging with a single asset class. Using a uniform price distribution and some select parameters, Morris proved that an investor who buys N shares over T time steps using dollar-cost averaging is expected to pay less per share than an investor who buys $\frac{N}{T}$ shares per time step. Wilson (1961) extended the notion of dollar-cost averaging given by Morris to accelerated dollar-cost averaging for a single asset class. Accelerated dollar-cost averaging makes purchases a function of the squared price in order to examine price sensitivity. This construction was created solely to examine the effect of price sensitivity on an abstract notion of dollar-cost averaging. Theoretical results hint that increased price sensitivity results in an even lower expected price per share.

In [7], Dolley presented the dollar-cost averaging strategy in a large-scale, multi-asset setting. Using dollar-cost averaging, the University of Texas Permanent University Fund was able to achieve roughly market performance without needing to rely so heavily on human judgement. In addition, Dolley argued that funds reliant on human judgement are likely not to beat the market in the long run because the good and bad judgements balance each other out.

Khouja and Lamb (1999) optimized the time step length and the amount invested at each time step for dollar-cost averaging with a single asset class. Important considerations like commissions and the ability to automate investments via direct deposit of paychecks were taken into account during optimization.

Dubil (2004) showed via simulation that, in the long-term, dollar-cost averaging offers superior risk and reliability benefits when compared to the lump sum strategy. In fact, long-term investment in high risk assets is supported by dollar-cost averaging because the risk reduction is so significant. Moreover, dollar-cost averaging is more reliable in that it is less likely to fall short of a retirement target.

Kirkby, Mitra and Nguyen (2020) rigorously developed the dollar-cost averaging strategy with one asset class, also incorporating an interest earning cash reserve from which funds are pulled to buy shares of the asset class. An exponential Levy process was used as framework, and the dollar-cost averaging strategy is compared with a lump sum strategy. The lump sum strategy offers better returns, but at higher risk. To describe this risk-return tradeoff in depth, Kirkby, Mitra and Nguyen introduced and rigorously developed a strategy that is part dollar-cost averaging and part lump sum. Balvers and Mitchell (1997) used a normal ARMA(1,1) process to examine dollar-cost averaging, also finding that gradual entry into an asset class is suboptimal.

Value averaging is a more recent development compared to dollar-cost averaging. In 1988, Michael Edleson introduced the notion of value averaging to contrast with dollar-cost averaging, see [10]. The original idea involved a linear target amount, one risky asset and a cash reserve. Using market data, value averaging was shown to outperform dollar-cost averaging in annual return.

Eng and Wang (2007) used simulation to compare dollar-cost averaging, value averaging and the lump sum strategy when prices are mean reverting. An exponential target amount was used, and results supported the idea that mean reverting prices favor value averaging.

Chen and Estes (2010) used Monte Carlo simulations to compare dollar-cost and value averaging with two asset classes: stocks and bonds. Value averaging was deemed the all-around winner in terms of the risk-return balance. No theoretical attempt was made to describe the distribution of returns for dollar-cost and value averaging.

The risk-return tradeoff of value averaging with linear target amounts has been analyzed empirically [14]. Sharpe ratio was used to measure risk-return tradeoff. Results indicated long-term implementation of value averaging produces higher Sharpe ratios.

Lai, Tseng and Huang (2016) introduced a modified value averaging strategy that uses Bollinger Bands to determine entrance and exit times. In simulation, the modified strategy outperformed traditional value averaging.

It is conceivable to extend value averaging to an arbitrary number of asset classes using multiple target amounts and conditions dictating where funds should go when target amounts are met or not. However this gets quite complicated for theoretical results, and research on value averaging has not ventured into this realm.

1.2 Goal of theoretical results

Thus far, dollar-cost averaging has been described rigorously with one asset class [16]. Value averaging has yet to be described rigorously. This thesis describes the dollar-cost and value averaging strategies for two asset classes via distribution, expected value and variance of returns. The idea is that, given two asset classes, an investor can assess the risk-return tradeoff between strategies and parameters using the theoretical distribution, expected value and variance of returns. Note that full realization of this idea depends on computation of the theoretical expressions developed for distribution, expected value and variance of returns. As it turns out, only the theoretical results having to do with expectated value and variance of dollar-cost averaging returns are feasible to compute.

Because dollar-cost averaging is more tractable than value averaging, the expected value, variance and Sharpe ratio are optimized as well. Note that Sharpe ratio is used to measure risk-return tradeoff, as it provides the ratio between expected value and standard deviation of returns [21]. The Sharpe ratio was selected here because it is a function of expected value and variance, which have closed form expression in the dollar-cost averaging case. A higher Sharpe ratio is supposed to indicate a better investment. However, the Sharpe ratio can be misleading in certain situations, like when the expected value is positive but the variance is so near to 0 that the Sharpe ratio is absurdly high. As a consequence, it is best to support the Sharpe ratio with additional information, like quantiles of returns.

1.3 Stocks and bonds

Stock represents part-ownership in a corporation. As a result of this part-ownership, stock owners recieve voting rights and a portion of a corporation's profits, namely dividends, depending on how much stock they own. Some corporations do not pay dividends, instead offering stock buybacks to maintain stock prices.

Stocks are sold at market value, which can be above or below the price initially payed for the stock. Volatility in the market makes stock investment risky. However, this added risk translates to increased return, making stocks attractive to investors wishing for high returns [18].

Bonds are nearly risk-free investments. They redeem at par value, which can be above or below the price initially payed for the bond [3]. Par value is what the issuer promises to pay bondholders when the bond matures. Furthermore, par value is set when the bond is initially purchased. The risk comes from probability of a default. Risk is avoided by choosing insured bonds or issuers with outstanding credit. Bonds can be withdrawn early for a penalty or traded in the secondary market.

The maturity on bonds can be short, like 1 month, or long, like 20 years. In general, longer maturities offer better returns. Bonds are often issued by banks and governments to raise funds. Unlike stocks, bonds have a promised return on maturity that makes bonds attractive to investors looking for positive returns at low risk.

Investors in stocks and bonds should be aware of the impact of brokers, the consumer price index and taxes on returns. Brokers like US Bank and Vanguard act as a medium through which investors buy and sell stock and bonds. They charge fees for their service as middlemen. If an investor reads the fine print and chooses a broker carefully, fees can be reduced to very small levels. The consumer price index tracks inflation in consumer goods and services. After accounting for inflation, a risk-free bond investment can actually yield a negative return if inflation outpaces the par value. Taxes must be paid on capital gains, which are realized by selling stock or redeeming bonds for profit. More details about taxes can be found in [18].

1.4 Index funds and diversification

An index fund is a collection of assets designed to summarize a segment of the financial market. Usually the assets in a given index fund are from the same class, like the class of stocks. Alternatively, the asset class can be more specific, like US stocks.

Index funds are attractive investments because of their diversification [18]. Diversification is defined here as low correlation between assets. If some assets in a fund lose significant value, the remaining assets prop the fund up. In effect, the investor feels a fraction of that loss. Because stocks are so risky, index funds are especially attractive. The diversification reduces risk and makes stocks palatable for investors. Through index funds, investors are able to get the increased returns offered by stocks at a manageable level of risk.

Here is the reason owning stock in a small number of corporations is so dangerous. Stocks represent individual corporations, each with unique characteristics. Even in the same industry, corporations have idiosyncrasies that can lead one into bankruptsy and the other into prosperity. It is extremely difficult to predict which corporations will tend toward bankruptsy vs prosperity. Moreover, the impact of dramatic loss is most significant when a large proportion of an investor's portfolio is stock with that loss. The diversification of index funds reduces the impact of a dramatic loss and helps investors avoid the possibility of holding worthless investments.

1.5 Goal of applications

The dollar-cost and value-averaging strategies are applied to a US stock index fund and short-term US bonds. More details on the particular US stock index and short-term US bonds can be found in Chapter 2. The idea behind choosing a high risk, high return stock index and low risk, low return bonds is that an investor, by selection of parameters, can achieve a desired risk-return tradeoff that is somewhere inbetween high risk, high return and low risk, low return. Furthermore, applications provide investors with several figures describing what the various risk-return tradoffs are for a range of parameters. When computation is feasible, figures are produced for theoretical results. Otherwise, simulation is used to produce figures. After presenting applications, an arguement is made in favor of dollar-cost averaging for large investment funds like university endowment funds, see Section 8.1.

1.6 Organization

The remainder of the thesis is organized as follows.

- Chapter 2 describes the data used in applications and the abreviations used to shorten notation.

- Chapter 3 describes the general methods, including notation, used throughout.

Methods pertaining to dollar-cost and value averaging are also described, but separately.

- Chapter 4 gives the theoretical results for dollar-cost averaging. First the cumulative distribution function for real returns is presented through recursion. Next the expectation and variance of real returns are given recursively and in closed-form. Last, the expectation, variance and Sharpe ratio of real returns are optimized.

- Chapter 5 gives the theoretical results for value averaging. First, relevant cumulative distribution functions are presented through recursion. Then related expectations and variances are described using those cumulative distribution functions.

- Chapter 6 investigates the distribution of real returns given by the data to be used in applications. Autoregression is fitted, and residuals are modeled with a Normal distribution and an alternate kernel density estimate for comparison.

- Chapter 7 first applies the dollar-cost averaging theoretical results and summarizes observations. Value averaging theoretical results are not applied because they take too much time and space to compute. Next the dollar-cost and value averaging strategies are simulated. Distributions of real returns are described with quantiles, expectation, standard deviation and Sharpe ratio. Observations about the distributions of real returns are summarized for each strategy.

- Chapter 8 provides an argument in favor of dollar-cost averaging for large investment funds. Then some final remarks are given.

- Chapter 9 is an appendix containing details used in the kernel density estimate from Chapter 6.

2. DATA & ABBREVIATIONS

2.1 Data

The stock and bond data used in applications describes 1871 to 2020 on an annual basis [5]. Relevant variables are described below.

Tab. 2.1: Data variable descriptions

Notation	Description
Index	average monthly close of the S&P composite index
Dividend	dividend per share of the S&P composite index
CPI	January consumer price index
1YTR	returns from 6-month bonds, averaged January and July

S&P and CPI data was taken from [22]. Specifically, the S&P indexes used were Cowles and Associates from 1871 to 1926, Standard & Poor 90 from 1926 to 1957 and Standard & Poor 500 from 1957 to 2020. The 1YTR data was taken from [22] for the period 1871-2011. Specifically, the 1YTR rates were off US commercial paper from 1871 to 1997 and US certificates of deposit, secondary market rate, from 1998 to 2011. The 1YTR data was taken from [20] for the period 2012-2019. Specifically, the 1YTR rates were off US certificates of deposit, non-jumbo deposits.

2.1.1 S&P Composite Index

There are three S&P composite indexes used in applications. All were developed to summarize the US stock market in terms of market capitalization [25]. Companies are given different weights in an index based on their individual market capitalization. The S&P 90 consists of 90 companies and the S&P 500 consists of 500 companies [12]. The Cowles and Associates index is a backward extension of the S&P index.

2.1.2 6-month bonds

The 6-month bonds consist of US commerical paper and US certificates of deposits. Both are short-term, low-risk instruments used by lenders to accrue interest on available capital [2]. Commercial paper is traded at a discount that accounts for current interest rates. Certificates of deposit are receipts for bank deposits, which accrue interest and cannot be withdrawn until a set maturity.

Note that the average of January and July 6-month bond rates is used instead of the 1 year bond rate because Robert Shiller and colleagues assembled a long-term spreadsheet of the former only [22]. Applications could have used 1 year bond rates, but the most reliable source was Shiller's website. Consequently, the average of 6-month bond rates is used in applications.

2.2 Abbreviations

In order to simplify statements, common phrases are abbreviated.

Tab. 2.2: Abbreviations

Notation	Description
S&P	S&P Composite Index
1YTR	6-month bonds
RR	Real returns
RR+1	Add 1 to real returns
DCA	Dollar-cost averaging
VA	Value averaging
DCRR	Dollar-cost averaging real returns
VARR	Value averaging real returns

3. METHODS

3.1 General methods

Both the dollar-cost and value averaging investing strategies begin with the following setup. Take two distinct asset classes, denoted A and B. If an asset class consists of multiple individual assets, bundle the desired individual assets into a single broad-based asset; this broad-based asset will be used to represent that asset class. Next, assume RR+1 of the two asset classes is a continuous bivariate random variable. Approximate the distribution of the RR+1 bivariate random variable and simulate n time steps. The n simulations are stored in a $n \times 2$ matrix X. The first column holds asset class A's RR+1 and the second column holds asset class B's RR+1. The ith row represents RR+1 for time step i.

In this paper, the two asset classes used are S&P and 1YTR with an annual time step. Asset class A is S&P and asset class B is 1YTR.

All random variables are assumed to be real-valued. When working with continuous and mixed random variables, the probability density functions are denoted with f and cumulative distribution functions are denoted with F. Variables involved are indicated by the subscript.

3.1.1 Real returns

Real return is defined as

$$\text{Real return } = \frac{1 + \text{ Nominal return}}{1 + \text{ Inflation rate}} - 1.$$

Now the actual (not simulated) annual RR of S&P and 1YTR is formulated for use in applications. Use the subscript k to denote the kth year of a given variable described in Section 2.1. Let s_k denote the RR during year k from S&P.

$$s_k = \frac{\frac{\textbf{Index}_{k+1}+\textbf{Dividend}_k}{\textbf{Index}_k}}{\frac{\textbf{CPI}_{k+1}}{\textbf{CPI}_k}} - 1 \tag{3.1}$$

Similarly, let b_k denote the RR during year k from 1YTR.

$$b_k = \frac{1 + \frac{\textbf{1YTR}_k}{100}}{\frac{\textbf{CPI}_{k+1}}{\textbf{CPI}_k}} - 1 \tag{3.2}$$

Figure 3.1 illustrates 149 years of annual RR, computed using (3.1) and (3.2). Observe the distinct change in variance of 1YTR real returns around 1950. This is a result of the Treasury-Fed Accord of 1951, which removed the connection between US monetary policy and management of government debt [13]. In order to account for this significant historical event, applications will split the data at 1951.

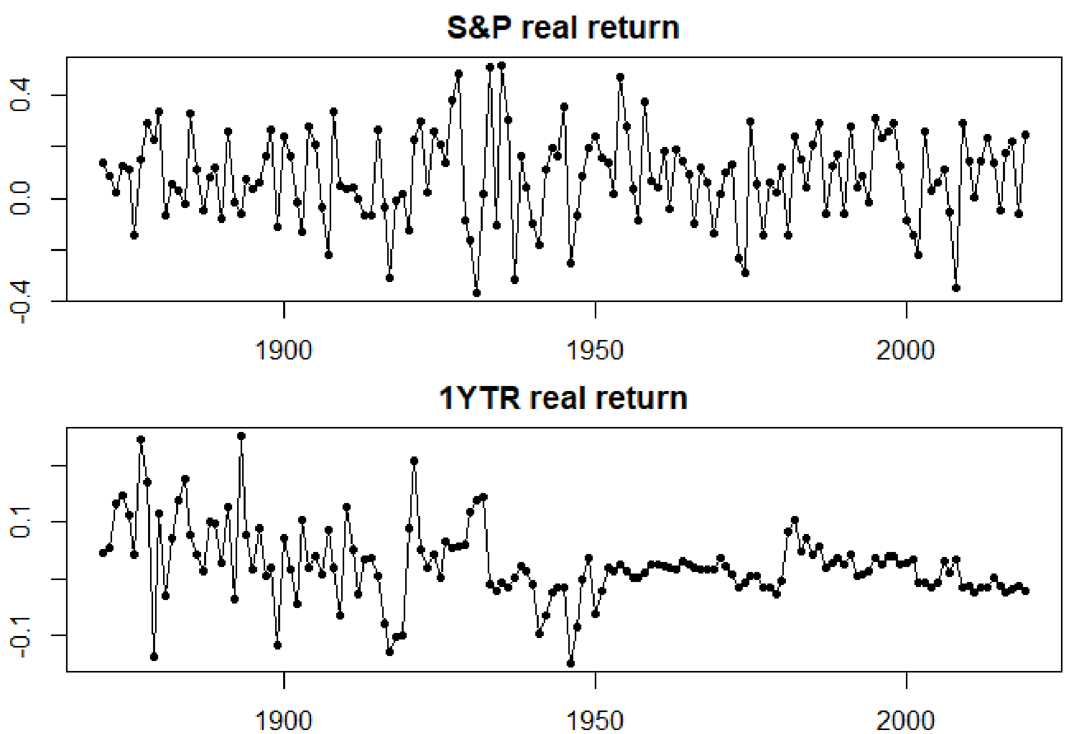

Fig. 3.1: Annual RR for S&P and 1YTR from 1871 to 2019

3.2 *Dollar-cost averaging methods*

In dollar-cost averaging, a constant amount is invested at each time step. When investing in two asset classes, a proportion of that constant amount is invested in each asset class. The algorithm below is used to compute the total RR after using dollar-cost averaging for n years.

Algorithm 1 DCRR after n years of value averaging

```
p ∈ [0,1] {Proportion invested in S&P}
TI ← 0 {Total invested}
VoBI ← 0 {Value of bond (1YTR) investments}
VoSI ← 0 {Value of stock (S&P) investments}
for i ← 1 to n do
    VoBI ← X_{i2}(VoBI + 1 − p)
    VoSI ← X_{i1}(VoSI + p)
end for
DCRR ← (VoSI+VoBI)/TI − 1
```

The recursion is described with subscripts below.

$$
\begin{aligned}
DC_{11} &= X_{11},\ DC_{12} = X_{12} \\
DC_{k1} &= X_{k1}\left(1 + DC_{k-1,1}\right) && 2 \le k \le n \\
DC_{k2} &= X_{k2}\left(1 + DC_{k-1,2}\right) && 2 \le k \le n \\
DCRR_k &= \frac{1}{k}(pDC_{k1} + (1-p)DC_{k2} - k) && 1 \le k \le n.
\end{aligned}
$$

Alternatively, $DCRR_n$ can be expressed via the following explicit formula. Define the $n \times 2$ matrix X' such that $X'_{ij} = \prod_{k=n-i+1}^{n} X_{kj}$. Then X'_{ij} represents the cumulative RR+1 from time step $n-i+1$ to n. Moreover, the total RR after n time steps of using dollar-cost averaging with $p \in [0,1]$ of the constant investment amount going to S&P and $1-p$ going to asset 1YTR is given by

$$DCRR_n = \frac{p}{n}\sum_{i=1}^{n} X'_{i1} + \frac{1-p}{n}\sum_{i=1}^{n} X'_{i2} - 1. \tag{3.3}$$

3.3 Value averaging methods

In value averaging, a target amount is set for the beginning of each year, and it is acheived by buying or selling assets. When investing in two assets, target amounts are placed on the riskier asset. The less risky asset is used as a reserve for when the target is surpassed. Here, S&P is the risky asset and 1YTR is the less risky asset. The target amount for S&P is increased exponentially with time by the function $(1+r)^{t-1}$ where r denotes the desired growth rate and t denotes the time in years. Every time S&P is sold to meet a target, the money from the sale is used to buy 1YTR. If S&P needs to be bought in order to meet a target, money for the purchase is pulled from 1YTR first; any additional money required is pulled from a bottomless cash reserve.

The algorithm below is used to compute the total RR after using value averaging for n years.

Algorithm 2 VARR after n years of value averaging

$T \leftarrow \frac{1}{1+r}$ {Target amount}
$TI \leftarrow 0$ {Total invested}
$VoBI \leftarrow 0$ {Value of bond (1YTR) investments}
$VoSI \leftarrow 0$ {Value of stock (S&P) investments}
for $i \leftarrow 1$ to n **do**
  $T \leftarrow (1+r) \cdot T$
  $VoBI \leftarrow VoBI + VoSI - T$
  if $VoBI < 0$ **then**
    $TI \leftarrow TI - VoBI$
    $VoBI \leftarrow 0$
  end if
  $VoSI \leftarrow X_{i1} \cdot T$
  $VoBI \leftarrow X_{i2} \cdot VoBI$
end for
$VARR \leftarrow \frac{VoSI+VoBI}{TI} - 1$

Now define the function $g : \mathbb{R} \to \mathbb{R}$ such that $g(x) = \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$. Using similar notation as in Algorithm 2, let the subscript k denote a variable's state at the end of year k, before the next year's investments are made. The recursion is defined with subscripts below where $1 \leq k \leq n$.

$$\begin{aligned}
0 &= VoSI_0 = VoBI_0 = TI_0 \\
T_k &= (1+r)^{k-1} \\
VoSI_k &= X_{k1}T_k \\
A_k &= VoBI_{k-1} + VoSI_{k-1} - T_k \\
VoBI_k &= X_{k2}g(A_k) \\
TI_k &= TI_{k-1} + g(-A_k) \\
VARR_k &= \frac{VoSI_k + VoBI_k}{TI_k} - 1.
\end{aligned}$$

4. DOLLAR-COST AVERAGING THEORETICAL RESULTS

Simplicity of the dollar-cost averaging recursion makes for straightforward derivation of the distribution, expectation and variance of $DCRR_n$. First the distribution, expectation and variance of $DCRR_n$ are found via recursion. Then the expectation and variance of $DCRR_n$ are described in closed-form.

4.1 $DCRR_n$ Cumulative distribution function

Theorem 1. *Let $2 \leq k \leq n$, $Z_1 = (X_{k1}, X_{k2})$ and $Z_2 = (DC_{k-1,1}, DC_{k-1,2})$. Then (DC_{k1}, DC_{k2}) and $DCRR_n$ are continuous,*

$$f_{DC_{11},DC_{12}}(\alpha,\beta) = f_{X_{11},X_{12}}(\alpha,\beta)$$
$$f_{DC_{k1},DC_{k2}}(\alpha,\beta) = \int_{(\mathbb{R}-\{0\})^2} f_{Z_1}(x_1,\ x_2) f_{Z_2}(\frac{\alpha}{x_1} - 1,\ \frac{\beta}{x_2} - 1)\frac{1}{|x_1 x_2|} dV(x_1, x_2).$$

and

$$F_{DCRR_n}(a) = \int_{-\infty}^{a} \int_{\mathbb{R}} f_{DC_{n1},DC_{n2}}(\alpha,\ \frac{1}{1-p}(n\beta + n - p\alpha))\frac{n}{1-p} d\alpha d\beta.$$

Proof. The continuity of (DC_{k1}, DC_{k2}) follows from linearity in its recursive definition and the continuity of (X_{j1}, X_{j2}) for all $1 \leq j \leq k$. Similarly, the continuity of $DCRR_n$ follows from its definition as a linear combination of the continuous DC_{n1} and DC_{n2}. The cdf of $DCRR_n$ is attained through recursion because the two sums of products in (3.3) make it too difficult to formulate a cdf otherwise.

First observe that $DCRR_n$ is a linear combination of DC_{n1} and DC_{n2}, so it suffices to find the joint cdf of DC_{n1} and DC_{n2} and then make a transformation to get the cdf of $DCRR_k$.

Define the transformation $G : (\mathbb{R} - \{0\})^2 \times \mathbb{R}^2 \to (\mathbb{R} - \{0\})^2 \times \mathbb{R}^2$ such that $G(x_1,\ x_2,\ \alpha,\ \beta) = (x_1,\ x_2,\ x_1(1 + \alpha),\ x_2(1 + \beta))$.

The transformation G is one-to-one because $G(a_1, a_2, a_3, a_4) = G(b_1, b_2, b_3, b_4)$ implies $(a_1, a_2, a_1(1 + a_3), a_2(1 + a_4)) = (b_1, b_2, b_1(1 + b_3), b_2(1 + b_4))$. Since $a_1 = b_1$ and $a_2 = b_2$, it follows from substitution that $a_3 = b_3$ and $a_4 = b_4$.

Let $(x_1,\ x_2,\ \alpha,\ \beta) \in (\mathbb{R} - \{0\})^2 \times \mathbb{R}^2$. The transformation G is onto because

$$\begin{aligned} G(x_1,\ x_2,\ \frac{\alpha}{x_1} - 1,\ \frac{\beta}{x_2} - 1) &= (x_1,\ x_2,\ x_1(1 + \frac{\alpha}{x_1} - 1),\ x_2(1 + \frac{\beta}{x_2} - 1)) \\ &= (x_1,\ x_2,\ \alpha,\ \beta). \end{aligned}$$

Since $x_1, x_2 \neq 0$, G^{-1} is clearly C^1 and the Jacobian of G^{-1} is given by

$$\left| \frac{\partial(x_1,\ x_2,\ \frac{\alpha}{x_1} - 1,\ \frac{\beta}{x_2} - 1)}{\partial(x_1,\ x_2,\ \alpha,\ \beta)} \right| = |det(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{-\alpha}{x_1^2} & 0 & \frac{1}{x_1} & 0 \\ 0 & \frac{-\beta}{x_2^2} & 0 & \frac{1}{x_2} \end{bmatrix})| = \frac{1}{|x_1 x_2|}.$$

Let $Y = \{(x_1,\ x_2,\ \alpha,\ \beta) \in \mathbb{R}^4 :\ x_1, x_2 \in \mathbb{R},\ \alpha \leq a,\ \beta \leq b\}$ and $Y' = Y - (0) \times \mathbb{R}^3 - \mathbb{R} \times (0) \times \mathbb{R}^2$. Then

$$\begin{aligned} &P(DC_{k1} \leq a,\ DC_{k2} \leq b) \\ &\qquad = P((X_{k1}, X_{k2}, DC_{k1}, DC_{k2}) \in Y) \\ &\qquad = P((X_{k1}, X_{k2}, DC_{k1}, DC_{k2}) \in Y') \\ &\qquad\quad + P((X_{k1}, X_{k2}, DC_{k1}, DC_{k2}) \in (0) \times \mathbb{R}^3) \\ &\qquad\quad + P((X_{k1}, X_{k2}, DC_{k1}, DC_{k2}) \in \mathbb{R} \times (0) \times \mathbb{R}^2) \\ &\qquad = P((X_{k1}, X_{k2}, DC_{k1}, DC_{k2}) \in Y'). \end{aligned}$$

G being one-to-one and onto allows

$$\begin{aligned} &P((X_{k1}, X_{k2}, DC_{k1}, DC_{k2}) \in Y') \\ &\qquad = P(G^{-1}(X_{k1}, X_{k2}, DC_{k1}, DC_{k2}) \in G^{-1}(Y')) \\ &\qquad = P((X_{k1}, X_{k2}, DC_{k-1,1}, DC_{k-1,2}) \in G^{-1}(Y')). \end{aligned} \tag{4.1}$$

In order to shorten notation, let $Z = (Z_1, Z_2)$ where $Z_1 = (X_{k1}, X_{k2})$ and $Z_2 = (DC_{k-1,1}, DC_{k-1,2})$. Observe that Z_2 is a sum of products of continuous random variables, so Z_2 is continuous. Moreover, Z_1 is continuous by definition. From the continuity of Z, the last expression in (4.1) can be written as

$$\int_{G^{-1}(Y')} f_Z(x_1,\ x_2,\ \alpha,\ \beta) dV(x_1,\ x_2,\ \alpha,\ \beta). \tag{4.2}$$

Using the change of variables described in [23], the integral in (4.2) is transformed into

$$\int_{Y'} f_Z(x_1,\ x_2,\ \frac{\alpha}{x_1}-1,\ \frac{\beta}{x_2}-1)\frac{1}{|x_1x_2|}dV(x_1,\ x_2,\ \alpha,\ \beta). \tag{4.3}$$

The independence of X_{ki} and $DC_{k-1,j}$ for $i,j=1,2$ allows (4.3) to be written as,

$$\int_{Y'} f_{Z_1}(x_1,\ x_2)f_{Z_2}(\frac{\alpha}{x_1}-1,\ \frac{\beta}{x_2}-1)\frac{1}{|x_1x_2|}dV(x_1,\ x_2,\ \alpha,\ \beta). \tag{4.4}$$

Taking the derivative of (4.4) with respect to α and β yields the pdf of $(DC_{k,1}, DC_{k,2})$.

$$f_{DC_{k,1},DC_{k,2}}(\alpha,\beta) = \int_{(\mathbb{R}-\{0\})^2} f_{Z_1}(x_1,\ x_2)f_{Z_2}(\frac{\alpha}{x_1}-1,\ \frac{\beta}{x_2}-1)\frac{1}{|x_1x_2|}dV(x_1,x_2). \tag{4.5}$$

So the $f_{DC_{k,1},DC_{k,2}}$ can be found via recursion as long as f_{Z_1} is known or estimated.

Next define the transformation $G:\mathbb{R}^2\to\mathbb{R}^2$ such that $G(\alpha,\ \beta)=(\alpha,\ \frac{1}{n}(p\alpha+(1-p)\beta-n))$.

The transformation G is clearly one-to-one and onto. Moreover, the inverse is given by $G^{-1}(\alpha,\ \beta)=(\alpha,\ \frac{1}{1-p}(n\beta+n-p\alpha))$. G^{-1} is clearly C^1 and the Jacobian of G^{-1} is

$$\left|\frac{\partial(\alpha,\ \frac{1}{1-p}(n\beta+n-p\alpha))}{\partial(\alpha,\ \beta)}\right| = |det(\begin{bmatrix} 1 & 0 \\ \frac{-p}{1-p} & \frac{n}{1-p}\end{bmatrix})| = \frac{n}{1-p}.$$

Let $Y=\mathbb{R}\times(-\infty,a]$. G being one-to-one and onto allows

$$\begin{aligned} P(DCRR_n\le a) &= P((DC_{n1}, DCRR_n)\in Y) \\ &= P(G^{-1}(DC_{n1}, DCRR_n)\in G^{-1}(Y)) \\ &= P((DC_{n1}, DC_{n2})\in G^{-1}(Y)). \end{aligned}$$

From the continuity of (DC_{n1}, DC_{n2}),

$$P((DC_{n1}, DC_{n2})\in G^{-1}(Y)) = \int_{G^{-1}(Y)} f_{DC_{n1},DC_{n2}}(\alpha,\ \beta)dV(\alpha,\ \beta). \tag{4.6}$$

Again using the change of variables described in [23], the integral in (4.6) is transformed into

$$\int_Y f_{DC_{n1},DC_{n2}}(\alpha,\ \frac{1}{1-p}(n\beta+n-p\alpha))\frac{n}{1-p}dV(\alpha,\ \beta).$$

□

4.2 DCRR Expectation and variance

Assuming bivariate real returns are independent and identically distributed, the expectation and variance of $DCRR_n$ can be formulated with and without recursion. The method with recursion is easier to achieve. Assuming autoregression in the real returns is reasonable, but it adds considerable complexity and interrupts application of the linear properties for expectation and variance. As a result, autoregression is not considered here. All theorems within this chapter use the assumption that bivariate real returns are independent and identically distributed. The mean and covariance of the iid $[X_{k1}, X_{k2}]^T$ will be notated as

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}. \tag{4.7}$$

The following lemma is needed for the formulation without recursion.

Lemma 1. *For* $j = 1, 2$, $E[X'_{ij}] = \mu_j^i$.

Proof. Substituting X'_{ij} with its product form yields $E[X'_{ij}] = E[\prod_{k=n-i+1}^{n} X_{kj}]$. Since the X_{kj} are independent, the expectation of products becomes a product of expectations.

$$E[\prod_{k=n-i+1}^{n} X_{kj}] = \prod_{k=n-i+1}^{n} E[X_{kj}]. \tag{4.8}$$

Substituting μ_j for $E[X_{kj}]$ in the righthand side of (4.8) yields $\prod_{k=n-i+1}^{n} \mu_j$, which is further simplified to μ_j^i. □

4.2.1 $DCRR_k$ Expectation and variance (with recursion)

Theorem 2. *The expectation and variance of $DCRR_k$ for $1 \leq k \leq n$ is given recursively via the following set of equations.*

$$\begin{aligned}
E[DC_{11}] &= \mu_1,\ \ E[DC_{12}] = \mu_2 \\
Var(DC_{11}) &= \sigma_{11},\ \ Var(DC_{12}) = \sigma_{22},\ \ Cov(DC_{11}, DC_{12}) = \sigma_{12} \\
E[DC_{k1}] &= \mu_1 \left(1 + E[DC_{k-1,1}]\right) \\
E[DC_{k2}] &= \mu_2 \left(1 + E[DC_{k-1,2}]\right) \\
Var(DC_{ki}) &= (\sigma_{ii} + \mu_i^2) Var(DC_{k-1,i}) + \sigma_{ii}(1 + E[DC_{k-1,i}])^2 \\
Cov(DC_{k1}, DC_{k2}) &= \sigma_{12}(1 + E[DC_{k-1,2}] + E[DC_{k-1,1}]) \\
&\quad + Cov(DC_{k-1,1}, DC_{k-1,2}) \\
E[DCRR_k] &= \frac{1}{k}(pE[DC_{k1}] + (1-p)E[DC_{k2}] - k) \\
Var(DCRR_k) &= \frac{1}{k^2}(p^2 Var(DC_{k1}) + (1-p)^2 Var(DC_{k2}) \\
&\quad + 2p(1-p)Cov(DC_{k1}, DC_{k2})).
\end{aligned} \tag{4.9}$$

Proof. First the recursive formula for expectation is presented. From the linearity of expectation and independence of X_{ki} with $DC_{k-1,i}$ for $i = 1, 2$,

$$\begin{aligned}
E[DC_{11}] &= \mu_1,\ \ E[DC_{12}] = \mu_2 \\
E[DC_{k1}] &= \mu_1 \left(1 + E[DC_{k-1,1}]\right) \\
E[DC_{k2}] &= \mu_2 \left(1 + E[DC_{k-1,2}]\right) \\
E[DCRR_k] &= \frac{1}{k}(pE[DC_{k1}] + (1-p)E[DC_{k2}] - k).
\end{aligned}$$

Now the recursive formula for variance is developed. The recursion for variance requires more effort to develop than the recursion for expectation. For $i = 1, 2$ and $1 \leq k \leq n$ the definition of variance allows

$$\begin{aligned}
Var(DC_{ki}) &= E[DC_{ki}^2] - E[DC_{ki}]^2 \\
&= E[X_{ki}^2(1 + DC_{k-1,i})^2] - E[DC_{ki}]^2.
\end{aligned} \tag{4.10}$$

First employing the independence of X_{ki} with $(1 + DC_{k-1,i})$ and then the definition of variance,

$$\begin{aligned} E[X_{ki}^2(1 + DC_{k-1,i})^2] &= E[X_{ki}^2]E[(1 + DC_{k-1,i})^2] \\ &= (\sigma_{ii} + \mu_i^2)(Var(DC_{k-1,i}) + (1 + E[DC_{k-1,i}])^2). \end{aligned} \tag{4.11}$$

Substituting (4.11) into (4.10) yields

$$Var(DC_{ki}) = (\sigma_{ii} + \mu_i^2)(Var(DC_{k-1,i}) + (1 + E[DC_{k-1,i}])^2) - E[DC_{ki}]^2. \tag{4.12}$$

(4.12) is simplified by substituting $\mu_i(1 + E[DC_{k-1,i}])$ for $E[DC_{ki}]$ and reducing.

$$Var(DC_{ki}) = (\sigma_{ii} + \mu_i^2)Var(DC_{k-1,i}) + \sigma_{ii}(1 + E[DC_{k-1,i}])^2.$$

Moving to covariance, the linearity property of covariance is applied, taking into account the independence of $DC_{k-1,i}$ with X_{k1} and X_{k2} for $i = 1, 2$,

$$\begin{aligned} Cov(DC_{k1}, DC_{k2}) &= Cov(X_{k1}(1 + DC_{k-1,1}), X_{k2}(1 + DC_{k-1,2})) \\ &= Cov(X_{k1}, X_{k2}) + Cov(X_{k1}, X_{k2}DC_{k-1,2}) \\ &\quad + Cov(X_{k2}, X_{k1}DC_{k-1,1}) + Cov(DC_{k-1,1}, DC_{k-1,2}) \\ &= Cov(X_{k1}, X_{k2}) + E[DC_{k-1,2}]Cov(X_{k1}, X_{k2}) \\ &\quad + E[DC_{k-1,1}]Cov(X_{k2}, X_{k1}) + Cov(DC_{k-1,1}, DC_{k-1,2}) \\ &= \sigma_{12}(1 + E[DC_{k-1,2}] + E[DC_{k-1,1}]) \\ &\quad + Cov(DC_{k-1,1}, DC_{k-1,2}). \end{aligned}$$

The final step of the recursion is found using the quasi-linearity of variance,

$$\begin{aligned} Var(DCRR_k) = \frac{1}{k^2}(p^2Var(DC_{k1}) &+ (1 - p)^2Var(DC_{k2}) \\ &+ 2p(1 - p)Cov(DC_{k1}, DC_{k2})). \end{aligned}$$

Collecting all the recursive equations used for expectation and variance gives the result. □

4.2.2 $DCRR_n$ Expectation (without recursion)

Theorem 3. *The expectation of $DCRR_n$ is given by*

$$E[DCRR_n] = \frac{1}{n}\left(p\frac{\mu_1(1 - \mu_1^n)}{1 - \mu_1} + (1 - p)\frac{\mu_2(1 - \mu_2^n)}{1 - \mu_2}\right) - 1. \tag{4.13}$$

Proof. Here a closed-form representation for the expectation of $DCRR_n$ is developed based on (3.3).

Using the linearity of expectation and the fact that $E[1] = 1$,

$$E[DCRR_n] = \frac{1}{n}\left(p\sum_{i=1}^{n} E[X'_{i1}] + (1-p)\sum_{i=1}^{n} E[X'_{i2}]\right) - 1. \tag{4.14}$$

It follows from Lemma 1 that (4.14) can be simplified to

$$E[DCRR_n] = \frac{1}{n}\left(p\sum_{i=1}^{n} \mu_1^i + (1-p)\sum_{i=1}^{n} \mu_2^i\right) - 1. \tag{4.15}$$

Noticing the sums in (4.15) are geometric sums,

$$\sum_{i=1}^{n} \mu_j^i = \frac{\mu_j(1-\mu_j^n)}{1-\mu_j}, \quad j = 1, 2. \tag{4.16}$$

Substituting (4.16) into (4.15) gives the result. □

4.2.3 $DCRR_n$ Variance (without recursion)

Theorem 4. *The variance of $DCRR_n$ is given by*

$$\begin{aligned} Var(DCRR_n) = \frac{1}{n^2}\Bigg(& p^2 Var(\sum_{i=1}^{n} X'_{i1}) + (1-p)^2 Var(\sum_{i=1}^{n} X'_{i2}) \\ & +2p(1-p)Cov(\sum_{i=1}^{n} X'_{i1}, \sum_{i=1}^{n} X'_{i2})\Bigg). \end{aligned} \tag{4.17}$$

The Var and Cov expressions in (4.17) are found using (4.18).

$$\begin{aligned} Cov(\sum_{i=1}^{n} X'_{ij_1}, \sum_{i=1}^{n} X'_{ij_2}) = \sum_{i_1=1}^{n}\Bigg(& \sum_{i_2=i_1}^{n} (\sigma_{j_1j_2} + \mu_{j_1}\mu_{j_2})^{i_1}\mu_{j_2}^{i_2-i_1} \\ & + \sum_{i_2=1}^{i_1} (\sigma_{j_1j_2} + \mu_{j_1}\mu_{j_2})^{i_2}\mu_{j_1}^{i_1-i_2} \\ & -(\sigma_{j_1j_2} + \mu_{j_1}\mu_{j_2})^{i_1} - \sum_{i_2=1}^{n} \mu_{j_1}^{i_1}\mu_{j_2}^{i_2}\Bigg). \end{aligned} \tag{4.18}$$

Proof. An explicit formula for the variance is trickier to find compared to the expectation. It is necessary to work out $Cov(\sum_{i=1}^{n} X'_{ij_1}, \sum_{i=1}^{n} X'_{ij_2})$ first, where $j_1, j_2 = 1, 2$.

First the covariance of sums is expanded into the double sum of covariances.

$$Cov(\sum_{i=1}^{n} X'_{ij_1}, \sum_{i=1}^{n} X'_{ij_2}) = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} Cov(X'_{i_1j_1}, X'_{i_2j_2}) \tag{4.19}$$

Substituting for $Cov(X'_{i_1j_1}, X'_{i_2j_2})$ in the right-hand side of (4.19) using the definition of covariance,

$$\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}(E[X'_{i_1j_1}X'_{i_2j_2}] - E[X'_{i_1j_1}]E[X'_{i_2j_2}]). \tag{4.20}$$

Distributing the sum in (4.20) and substituting for $E[X'_{i_1j_1}]E[X'_{i_2j_2}]$ using Lemma 1,

$$\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E[X'_{i_1j_1}X'_{i_2j_2}] - \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \mu_{j_1}^{i_1}\mu_{j_2}^{i_2}. \tag{4.21}$$

Substituting $X'_{i_1j_1}$ and $X'_{i_2j_2}$ in (4.21) with their product forms,

$$\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E[\prod_{k=n-i_1+1}^{n} X_{kj_1} \prod_{k=n-i_2+1}^{n} X_{kj_2}] - \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \mu_{j_1}^{i_1}\mu_{j_2}^{i_2}. \tag{4.22}$$

Expanding the left-hand sum of (4.22),

$$\begin{aligned}
&\sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n} E[\prod_{k=n-i_1+1}^{n} X_{kj_1} \prod_{k=n-i_2+1}^{n} X_{kj_2}] \\
&+ \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1} E[\prod_{k=n-i_1+1}^{n} X_{kj_1} \prod_{k=n-i_2+1}^{n} X_{kj_2}] \\
&- \sum_{i_1=1}^{n} E[\prod_{k=n-i_1+1}^{n} X_{kj_1}X_{kj_2}] - \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \mu_{j_1}^{i_1}\mu_{j_2}^{i_2}.
\end{aligned} \tag{4.23}$$

Rearranging the products of (4.23),

$$\begin{aligned}
&\sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n} E[\prod_{k=n-i_1+1}^{n} X_{kj_1}X_{kj_2} \prod_{k=n-i_2+1}^{n-i_1} X_{kj_2}] \\
&+ \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1} E[\prod_{k=n-i_2+1}^{n} X_{kj_1}X_{kj_2} \prod_{k=n-i_1+1}^{n-i_2} X_{kj_1}] \\
&- \sum_{i_1=1}^{n} E[\prod_{k=n-i_1+1}^{n} X_{kj_1}X_{kj_2}] - \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \mu_{j_1}^{i_1}\mu_{j_2}^{i_2}.
\end{aligned} \tag{4.24}$$

Observing the independence of $[X_{kj_1}, X_{kj_2}]^T$ across k allows the expectation of products in (4.24) to be written as the product of expectations.

$$\begin{aligned}
&\sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n}\prod_{k=n-i_1+1}^{n} E[X_{kj_1}X_{kj_2}]\prod_{k=n-i_2+1}^{n-i_1} E[X_{kj_2}] \\
&+\sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1}\prod_{k=n-i_2+1}^{n} E[X_{kj_1}X_{kj_2}]\prod_{k=n-i_1+1}^{n-i_2} E[X_{kj_1}] \\
&-\sum_{i_1=1}^{n}\prod_{k=n-i_1+1}^{n} E[X_{kj_1}X_{kj_2}] - \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\mu_{j_1}^{i_1}\mu_{j_2}^{i_2}
\end{aligned} \tag{4.25}$$

From (4.25), substitute for $E[X_{kj_1}X_{kj_2}]$ using the definition of covariance. In addition, substitute μ_{j_1} and μ_{j_2} for $E[X_{kj_1}]$ and $E[X_{kj_2}]$.

$$\begin{aligned}
&\sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n}\prod_{k=n-i_1+1}^{n} (\sigma_{j_1j_2}+\mu_{j_1}\mu_{j_2})\prod_{k=n-i_2+1}^{n-i_1} \mu_{j_2} \\
&+\sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1}\prod_{k=n-i_2+1}^{n} (\sigma_{j_1j_2}+\mu_{j_1}\mu_{j_2})\prod_{k=n-i_1+1}^{n-i_2} \mu_{j_1} \\
&-\sum_{i_1=1}^{n}\prod_{k=n-i_1+1}^{n} (\sigma_{j_1j_2}+\mu_{j_1}\mu_{j_2}) - \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\mu_{j_1}^{i_1}\mu_{j_2}^{i_2}
\end{aligned} \tag{4.26}$$

Rewriting the products in (4.26) with exponents,

$$\begin{aligned}
&\sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n}(\sigma_{j_1j_2}+\mu_{j_1}\mu_{j_2})^{i_1}\mu_{j_2}^{i_2-i_1} + \sum_{i_1=1}^{n}\sum_{i_2=1}^{i_1}(\sigma_{j_1j_2}+\mu_{j_1}\mu_{j_2})^{i_2}\mu_{j_1}^{i_1-i_2} \\
&-\sum_{i_1=1}^{n}(\sigma_{j_1j_2}+\mu_{j_1}\mu_{j_2})^{i_1} - \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\mu_{j_1}^{i_1}\mu_{j_2}^{i_2}.
\end{aligned} \tag{4.27}$$

Factoring out the $\sum_{i_1=1}^{n}$ in (4.27) gives the finished equation for covariance,

$$\begin{aligned}
Cov\Big(\sum_{i=1}^{n}X'_{ij_1}, \sum_{i=1}^{n}X'_{ij_2}\Big) = \sum_{i_1=1}^{n}\Bigg(&\sum_{i_2=i_1}^{n}(\sigma_{j_1j_2}+\mu_{j_1}\mu_{j_2})^{i_1}\mu_{j_2}^{i_2-i_1} \\
&+\sum_{i_2=1}^{i_1}(\sigma_{j_1j_2}+\mu_{j_1}\mu_{j_2})^{i_2}\mu_{j_1}^{i_1-i_2} \\
&-(\sigma_{j_1j_2}+\mu_{j_1}\mu_{j_2})^{i_1} - \sum_{i_2=1}^{n}\mu_{j_1}^{i_1}\mu_{j_2}^{i_2}\Bigg).
\end{aligned}$$

From the quasi-linearity of variance, the variance of (3.3) is given by

$$Var(DCRR_n) = \frac{1}{n^2}\left(p^2 Var(\sum_{i=1}^{n} X'_{i1}) + (1-p)^2 Var(\sum_{i=1}^{n} X'_{i2}) \right.$$
$$\left. +2p(1-p)Cov(\sum_{i=1}^{n} X'_{i1}, \sum_{i=1}^{n} X'_{i2})\right).$$

□

4.3 $DCRR_n$ Optimization

The expectation, variance and Sharpe ratio of $DCRR_n$ are optimized over p.

4.3.1 $DCRR_n$ Argmax expectation

Theorem 5. *The argmax of $DCRR_n$ expectation over $p \in [0,1]$ is given by*

$$\arg\max_{p\in[0,1]} E[DCRR_n] = \begin{cases} 1 & \text{if} \quad \mu_1 \geq \mu_2 \\ 0 & \text{otherwise} \end{cases}.$$

Proof. First $E[DCRR_n]$ is differentiated. Using the explicit formula in (4.13) to take the derivative,

$$\frac{d}{dp}E[DCRR_n] = \frac{1}{n}\left(\frac{\mu_1(1-\mu_1^n)}{1-\mu_1} - \frac{\mu_2(1-\mu_2^n)}{1-\mu_2}\right). \tag{4.28}$$

Observing that $\frac{\mu_i(1-\mu_i^n)}{1-\mu_i} = \sum_{k=1}^{n} \mu_i^k$ for $i = 1, 2$, it is clear that $\frac{d}{dp}E[DCRR_n] \odot 0$ if and only if $\mu_1 \odot \mu_2$ where $\odot \in \{>, <, =\}$. If $\mu_1 > \mu_2$, then $\frac{d}{dp}E[DCRR_n] > 0$ on $p \in [0,1]$. Therefore $\max_{p\in[0,1]} E[DCRR_n] = E[DCRR_n]|_{p=1}$. Similarly, $\mu_1 < \mu_2$ implies $\max_{p\in[0,1]} E[DCRR_n] = E[DCRR_n]|_{p=0}$. When $\mu_1 = \mu_2$, $E[DCRR_n]$ is constant over $p \in [0,1]$. □

4.3.2 $DCRR_n$ Argmin variance

Theorem 6. *The argmin of $DCRR_n$ variance over p is given by*

$$\arg\min_p Var(DCRR_n)$$
$$= \frac{Var(\sum_{i=1}^{n} X'_{i2}) - Cov(\sum_{i=1}^{n} X'_{i1}, \sum_{i=1}^{n} X'_{i2})}{Var(\sum_{i=1}^{n} X'_{i1}) + Var(\sum_{i=1}^{n} X'_{i2}) - 2Cov(\sum_{i=1}^{n} X'_{i1}, \sum_{i=1}^{n} X'_{i2})}. \tag{4.29}$$

Proof. Using the explicit formula in (4.17) to take the derivative,

$$\frac{d}{dp}Var(DCRR_n) = \frac{2}{n^2}\Bigg(pVar(\sum_{i=1}^{n} X'_{i1}) - (1-p)Var(\sum_{i=1}^{n} X'_{i2}) \\ +(1-2p)Cov(\sum_{i=1}^{n} X'_{i1}, \sum_{i=1}^{n} X'_{i2})\Bigg). \tag{4.30}$$

Setting $\frac{d}{dp}Var(DCRR_n) = 0$ and using (4.30) to solve for p,

$$p = \frac{Var(\sum_{i=1}^{n} X'_{i2}) - Cov(\sum_{i=1}^{n} X'_{i1}, \sum_{i=1}^{n} X'_{i2})}{Var(\sum_{i=1}^{n} X'_{i1}) + Var(\sum_{i=1}^{n} X'_{i2}) - 2Cov(\sum_{i=1}^{n} X'_{i1}, \sum_{i=1}^{n} X'_{i2})}. \tag{4.31}$$

Furthermore, the second derivative of $Var(DCRR_n)$ with respect to p is

$$\frac{d^2}{dp^2}Var(DCRR_n) = \frac{2}{n^2}\Bigg(Var(\sum_{i=1}^{n} X'_{i1}) + Var(\sum_{i=1}^{n} X'_{i2}) \\ -2Cov(\sum_{i=1}^{n} X'_{i1}, \sum_{i=1}^{n} X'_{i2})\Bigg). \tag{4.32}$$

The right-hand side of (4.32) reduces to $\frac{2}{n^2}Var(\sum_{i=1}^{n} X'_{i1} - \sum_{i=1}^{n} X'_{i2})$, and it becomes clear that $\frac{d^2}{dp^2}Var(DCRR_n) \geq 0$. Therefore the solution given by (4.31) is indeed the value for p at which the minimum of $Var(DCRR_n)$ occurs. □

4.3.3 $DCRR_n$ Argmax Sharpe ratio

Theorem 7. *The derivative of the $DCRR_n$ Sharpe ratio $\frac{d}{dp}\frac{E[DCRR_n]}{\sqrt{Var(DCRR_n)}}$ is zero at*

$$p^* = -\frac{n(c - v_2) - bc + av_2}{v_1(n-b) + n(v_2 - 2c) + c(a+b) - av_2}.$$

where

$$a = E[\sum_{i=1}^{n} X'_{i1}] \qquad b = E[\sum_{i=1}^{n} X'_{i2}]$$

$$v_1 = Var(\sum_{i=1}^{n} X'_{i1}) \qquad v_2 = Var(\sum_{i=1}^{n} X'_{i2})$$

$$c = Cov(\sum_{i=1}^{n} X'_{i1}, \sum_{i=1}^{n} X'_{i2}).$$

Moreover, p^* is the argmax of the Sharpe ratio $\frac{E[DCRR_n]}{\sqrt{Var(DCRR_n)}}$ when $n \leq 50$, $1.11 > \mu_1 > \mu_2 > 1$ and $\sigma_{ii} > \sigma_{12}$ for $i = 1, 2$.

Proof. The Sharpe ratio $\frac{E[DCRR_n]}{\sqrt{Var(DCRR_n)}}$ is optimized. Taking the derivative via the quotient rule,

$$\frac{d}{dp}\frac{E[DCRR_n]}{\sqrt{Var(DCRR_n)}} = \frac{\sqrt{Var(DCRR_n)}\frac{d}{dp}E[DCRR_n] - E[DCRR_n]\frac{d}{dp}\sqrt{Var(DCRR_n)}}{Var(DCRR_n)}. \tag{4.33}$$

So $\frac{d}{dp}\frac{E[DCRR_n]}{\sqrt{Var(DCRR_n)}} = 0$ if and only if the numerator in the right-hand side of (4.33) is 0. Next, the numerator is rewritten using the chain rule.

$$\begin{aligned}
&\sqrt{Var(DCRR_n)}\frac{d}{dp}E[DCRR_n] - E[DCRR_n]\frac{d}{dp}\sqrt{Var(DCRR_n)} \\
&= \sqrt{Var(DCRR_n)}\frac{d}{dp}E[DCRR_n] - \frac{E[DCRR_n]\frac{d}{dp}Var(DCRR_n)}{2\sqrt{Var(DCRR_n)}} \\
&= \frac{1}{2\sqrt{Var(DCRR_n)}}\Big(2Var(DCRR_n)\frac{d}{dp}E[DCRR_n] \\
&\qquad -E[DCRR_n]\frac{d}{dp}Var(DCRR_n)\Big)
\end{aligned} \tag{4.34}$$

Using the last expression in (4.34), $\frac{d}{dp}\frac{E[DCRR_n]}{\sqrt{Var(DCRR_n)}} = 0$ when

$$2Var(DCRR_n)\frac{d}{dp}E[DCRR_n] - E[DCRR_n]\frac{d}{dp}Var(DCRR_n) = 0. \tag{4.35}$$

Substituting from (4.14), (4.17), (4.28) and (4.30), the solution of (4.35) is

$$p^* = -\frac{n(c - v_2) - bc + av_2}{v_1(n-b) + n(v_2 - 2c) + c(a+b) - av_2}. \tag{4.36}$$

where

$$a = E[\sum_{i=1}^{n} X'_{i1}] \qquad b = E[\sum_{i=1}^{n} X'_{i2}]$$

$$v_1 = Var(\sum_{i=1}^{n} X'_{i1}) \qquad v_2 = Var(\sum_{i=1}^{n} X'_{i2})$$

$$c = Cov(\sum_{i=1}^{n} X'_{i1}, \sum_{i=1}^{n} X'_{i2}).$$

But under what conditions is p^* the argmax of the Sharpe ratio? It suffices to look at the sign of the derivative on either side of p^*. From (4.33) and (4.34),

$\frac{d}{dp}\frac{E[DCRR_n]}{\sqrt{Var(DCRR_n)}}$ expands to

$$\frac{\sqrt{Var(DCRR_n)}\frac{d}{dp}E[DCRR_n] - E[DCRR_n]\frac{d}{dp}\sqrt{Var(DCRR_n)}}{2Var(DCRR_n)^{\frac{3}{2}}}. \tag{4.37}$$

The denominator of (4.37) is positive by the definition of variance and randomness of $Var(DCRR_n)$. Therefore, the sign of $\frac{d}{dp}\frac{E[DCRR_n]}{\sqrt{Var(DCRR_n)}}$ depends only on the numerator of (4.37). Consequently, p^* is the argmax if the derivative of the numerator is negative at p^*. Using the same notation that defined p^*, the derivative of the numerator of (4.37) with respect to p is

$$\frac{2}{n^3}\left((a-n)(c-v_2) + (b-n)(c-v_1)\right). \tag{4.38}$$

p^* can be verified as the argmax by checking (4.38) for negativity. Fortunately, this check can be avoided by fixing the parameter space to cover most likely investing scinerios. Assume that $n \leq 50$, $1.11 > \mu_1 > \mu_2 > 1$ and $\sigma_{ii} > \sigma_{12}$ for $i = 1, 2$. From the definitions of a and b, it follows that $a > b > n$. Therefore $a - n$ and $b - n$ are positive. So in order to show (4.38) is negative, it suffices to show that $c - v_i$ is negative for $i = 1, 2$. This is done through recursion.

Using the recursive definitions in (4.9), $Cov(DC_{k1}, DC_{k2}) - Var(DC_{ki})$ is expanded into

$$\begin{aligned} &\sigma_{12}(1 + E[DC_{k-1,2}] + E[DC_{k-1,1}]) + Cov(DC_{k-1,1}, DC_{k-1,2}) \\ &- (\sigma_{ii} + \mu_i^2)Var(DC_{k-1,i}) - \sigma_{ii}(1 + E[DC_{k-1,i}])^2. \end{aligned} \tag{4.39}$$

A short lemma is introduced to facilitate induction.

Lemma 2. *$\sigma_{ii}(1 + E[DC_{ki}])^2 > \sigma_{12}(1 + E[DC_{k2}] + E[DC_{k1}])$ for $i = 1, 2$ and $1 \leq k \leq n$, given $1.11 > \mu_1 > \mu_2 > 1$, $\sigma_{ii} > \sigma_{12}$ and $n \leq 50$.*

Proof. Since $\sigma_{ii} > \sigma_{12}$, it suffices to show that $(1+E[DC_{ki}])^2 > 1+E[DC_{k2}]+E[DC_{k1}]$. Furthermore, the assumption $\mu_1 > \mu_2$ implies $E[DC_{k1}] > E[DC_{k2}]$, so it suffices to show the latter with $i = 2$. From the definition of DC_{k2}, $\mu_2 > 1$ and $n \leq 50$ implies

$$E[DC_{k2}] + E[DC_{k2}]^2 > 50 + 50^2 = 2550. \tag{4.40}$$

Similarly, $\mu_1 < 1.11$ implies

$$E[DC_{k1}] < 1.11\frac{1-1.11^{50}}{1-1.11} = 1852.336. \tag{4.41}$$

After expanding and reducing, there is $(1+E[DC_{k2}])^2 > 1+E[DC_{k2}]+E[DC_{k1}]$ if and only if $E[DC_{k2}]+E[DC_{k2}]^2 > E[DC_{k1}]$. Moreover, (4.40) and (4.41) implies $E[DC_{k2}]+E[DC_{k2}]^2 > E[DC_{k1}]$. □

Now the induction begins. By assumption and definition of DC_{1i},

$$Cov(DC_{11}, DC_{12}) - Var(DC_{1i}) = \sigma_{12} - \sigma_{ii} < 0.$$

Suppose $Cov(DC_{k-1,1}, DC_{k-1,2}) - Var(DC_{k-1,i}) < 0$ for some $k \in \{2,..,n\}$. Since $\sigma_{ii} + \mu_i^2 > 1$, it follows that $Cov(DC_{k-1,1}, DC_{k-1,2}) - (\sigma_{ii}+\mu_i^2)Var(DC_{k-1,i}) < 0$. Lemma 2 implies $\sigma_{12}(1+E[DC_{k-1,2}]+E[DC_{k-1,1}]) - \sigma_{ii}(1+E[DC_{k-1,i}])^2 < 0$. The previous two statements and (4.39) imply $Cov(DC_{k1}, DC_{k2}) - Var(DC_{ki}) < 0$. The induction is finished.

In summary, the argmax of the Sharpe ratio is found using (4.36) and then checking (4.38) for negativity. In an effort to avoid the check for negativity, p^* is the argmax under the assumptions $n \leq 50$, $1.11 > \mu_1 > \mu_2 > 1$ and $\sigma_{ii} > \sigma_{12}$. Those assumptions fit realistic investing scinerios, so the check can likely be avoided in practice. □

5. VALUE AVERAGING THEORETICAL RESULTS

The decision over whether to move money out of or into 1YTR adds considerable complexity to value averaging - far more than in dollar-cost averaging. This added complexity limits theoretical results for value-averaging. Finding the distribution, expectation and variance of $VARR_n$ and TI_n is difficult, but it can be done through recursion. Note that in dollar-cost averaging, the total invested after n years is always $\frac{n(n+1)}{2}$. Here, the total invested is more complicated, so it is treated separately.

The main inroad into the complexity of value-averaging lies in A_k. First the distribution of A_k is found recursively. From there, it is straigtforward to formulate the distribution of $(TI_k,\ A_k)$. The distribution of $VARR_n$ is derived using the distribution of $(TI_k,\ A_k)$. Furthermore, expectations and variances are formulated from those distributions.

5.1 A_k Cumulative distribution function

The distribution of A_k is found via recursion. A lemma is provided as setup.

Lemma 3. *Let X, Y and Z be real-valued random variables such that Y is independent from X and Z, but X and Z are dependent. Moreover, X, Y and Z are continuous. Then $A = Xg(Y) + Z$ is continuous and has cdf and pdf given by*

$$F_A(a) = \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_0^{\infty} \frac{1}{y} f_{X,Z}(\frac{\alpha}{y}, z) f_Y(y) dy d\alpha dz + F_Z(a)F_Y(0)$$

$$f_A(a) = \int_{\mathbb{R}} \int_0^{\infty} \frac{1}{y} f_{X,Z}(\frac{a-z}{y}, z) f_Y(y) dy dz + f_Z(a)F_Y(0).$$

Proof. The definition of g allows A to be split into cases.

$$A = \begin{cases} XY + Z & \text{if} \quad Y > 0 \\ Z & \text{otherwize} \end{cases} \tag{5.1}$$

It follows from (5.1) and the Law of Total Probability that

$$P(A \leq a) = P(XY + Z \leq a,\ Y > 0) + P(Z \leq a,\ Y \leq 0). \tag{5.2}$$

Using the continuity of X, Y and Z along with the independence of Y and Z, (5.2) is rewritten as

$$P(A \leq a) = \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_{0}^{\infty} f_{XY,Y,Z}(\alpha, y, z) dy d\alpha dz + F_Z(a) F_Y(0). \tag{5.3}$$

The integral in (5.3) is transformed via the same method as in Section 4.1. Then the independence of Y with X and Z is applied.

$$\begin{aligned} P(A \leq a) &= \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_{0}^{\infty} \frac{1}{y} f_{X,Y,Z}(\frac{\alpha}{y}, y, z) dy d\alpha dz + F_Z(a) F_Y(0) \\ &= \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_{0}^{\infty} \frac{1}{y} f_{X,Z}(\frac{\alpha}{y}, z) f_Y(y) dy d\alpha dz + F_Z(a) F_Y(0). \end{aligned}$$

Lastly, observe that $P(A \leq a)$ is differentiable in a using Leibniz's integral rule and

$$\begin{aligned} \frac{d}{da} P(A \leq a) &= \frac{d}{da} \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_{0}^{\infty} \frac{1}{y} f_{X,Z}(\frac{\alpha}{y}, z) f_Y(y) dy d\alpha dz + \frac{d}{da} F_Z(a) F_Y(0) \\ &= \int_{\mathbb{R}} \frac{d}{da} \int_{-\infty}^{a-z} \int_{0}^{\infty} \frac{1}{y} f_{X,Z}(\frac{\alpha}{y}, z) f_Y(y) dy d\alpha dz + f_Z(a) F_Y(0) \\ &= \int_{\mathbb{R}} \int_{0}^{\infty} \frac{1}{y} f_{X,Z}(\frac{a-z}{y}, z) f_Y(y) dy dz + f_Z(a) F_Y(0). \end{aligned}$$

The differentiability of $P(A \leq a)$ implies A is continuous. □

Theorem 8. *Let $3 \leq k \leq n$. A_k is continuous with distribution given by*

$$\begin{aligned} F_{A_2}(a) &= \int_{-\infty}^{a} \int_{\mathbb{R}} \frac{1}{|T_1|} f_{X_{11},X_{12}}(\frac{z+T_2}{T_1}, x) dx dz \\ f_{A_2}(a) &= \int_{\mathbb{R}} \frac{1}{|T_1|} f_{X_{11},X_{12}}(\frac{a+T_2}{T_1}, x) dx \\ F_{A_k}(a) &= \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_{0}^{\infty} \frac{1}{y} \frac{1}{|T_{k-1}|} f_{X_{k-1,1},X_{k-1,2}} \left(\frac{z+T_k}{T_{k-1}}, \frac{\alpha}{y} \right) f_{A_{k-1}}(y) dy d\alpha dz \\ &\quad + \left(\int_{\mathbb{R}} \int_{-\infty}^{a} \frac{1}{|T_{k-1}|} f_{X_{k-1,1},X_{k-1,2}}(\frac{z+T_k}{T_{k-1}}, x) dz dx \right) F_{A_{k-1}}(0) \\ f_{A_k}(a) &= \int_{\mathbb{R}} \int_{0}^{\infty} \frac{1}{y} \frac{1}{|T_{k-1}|} f_{X_{k-1,1},X_{k-1,2}} \left(\frac{z+T_k}{T_{k-1}}, \frac{a-z}{y} \right) f_{A_{k-1}}(y) dy dz \\ &\quad + \left(\int_{\mathbb{R}} \frac{1}{|T_{k-1}|} f_{X_{k-1,1},X_{k-1,2}}(\frac{a+T_k}{T_{k-1}}, x) dx \right) F_{A_{k-1}}(0). \end{aligned}$$

Proof. By definition,

$$\begin{aligned} A_k &= VoBI_{k-1} + VoSI_{k-1} - T_k \\ &= X_{k-1,2}g(A_{k-1}) + X_{k-1,1}T_{k-1} - T_k. \end{aligned} \tag{5.4}$$

Now set $X = X_{k-1,2}$, $Y = A_{k-1}$ and $Z = X_{k-1,1}T_{k-1} - T_k$. The only dependent variables in the righthand side of (5.4) are $X_{k-1,2}$ and $X_{k-1,1}$, so it follows that $\text{Cov}(X, Y) = \text{Cov}(Y, Z) = 0$ and $\text{Cov}(X, Z) \neq 0$. Furthermore, the cdf $F_{X,Z}$ is found by making an integral transformation involving $f_{X_{k-1,1},X_{k-1,2}}$. Using the same method as in Section 4.1,

$$F_{X,Z}(x', z') = \int_{-\infty}^{x'} \int_{-\infty}^{z'} \frac{1}{|T_{k-1}|} f_{X_{k-1,1},X_{k-1,2}}(\frac{z + T_k}{T_{k-1}}, x)dzdx. \tag{5.5}$$

Differentiating (5.5) with respect to x and z yields the pdf

$$f_{X,Z}(x, z) = \frac{1}{|T_{k-1}|} f_{X_{k-1,1},X_{k-1,2}}(\frac{z + T_k}{T_{k-1}}, x). \tag{5.6}$$

The cdf of Z is found by taking a limit in (5.5).

$$F_Z(z) = \int_{\mathbb{R}} \int_{-\infty}^{z'} \frac{1}{|T_{k-1}|} f_{X_{k-1,1},X_{k-1,2}}(\frac{z + T_k}{T_{k-1}}, x)dzdx. \tag{5.7}$$

Observe that since $A_2 = X_{11}T_1 - T_2$, substituting $k = 2$ into (5.6) and (5.7) gives $f_{A_2}(a)$ and $F_{A_2}(a)$. Furthermore, A_2 is continuous by the continuity of X_{11}. This statement about the continuity of A_2 is necessary to start the induction showing the continuity of A_k.

Suppose Y is continuous for some $k \in \{2, .., n\}$. Since X and Z are continuous for all $k \in \{2, .., n\}$, Lemma 3 implies A_k is continuous. This completes the induction.

All conditions of Lemma 3 have been satisfied, so A_k is continuous and substitution from Lemma 3 gives $f_{A_k}(a)$ and $F_{A_k}(a)$. □

5.2 TI_k Cumulative distribution function

The procedure used to construct the distribution of A_k is modified to construct the distribution of (TI_k, A_k). A lemma is provided as setup.

Lemma 4. *Let W, X, Y and Z be real-valued random variables such that W and Y are independent from X and Z, but X and Z are dependent and W and Y are dependent. Moreover, W is mixed and X, Y and Z are continuous. The only discontinuity in W's cdf is at w^* and $F_W(w) = 0$ for $w < w^*$. Then (A, B), where $A = Xg(Y) + Z$ and $B = W + g(-Y)$, is mixed with continuous joint cdf except at $B = w^*$. Furthermore, $F_{A,B}(a, b) = 0$ for $b < w^*$ and*

$$
\begin{aligned}
F_{A,B}(a,b) = P(w^* \leq b) &\int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_0^{\infty} \frac{1}{y} f_{W,Y}(w^*, y) f_{X,Z}(\frac{\alpha}{y}, z) dy d\alpha dz \\
&+ \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_0^{\infty} \int_{w^*}^{b} \frac{1}{y} f_{W,Y}(w, y) f_{X,Z}(\frac{\alpha}{y}, z) dw dy d\alpha dz \\
&+ F_Z(a) \int_{w^*-b}^{0} f_{W,Y}(w^*, y) dy \\
&+ F_Z(a) \int_{-\infty}^{0} \int_{w^*}^{b+y} f_{W,Y}(w, y) dw dy
\end{aligned}
$$

$$
f_{A,B}(a,b) = \begin{cases} \frac{d^2}{dadb} F_{A,B}(a,b), & b \neq w^* \\ \frac{d}{da} F_{A,B}(a,b), & b = w^* \end{cases} .
$$

Proof. The definition of g allows A and B to be split into cases.

$$
A = \begin{cases} XY + Z & \text{if} \quad Y > 0 \\ Z & \text{otherwize} \end{cases} , \qquad B = \begin{cases} W & \text{if} \quad Y > 0 \\ W - Y & \text{otherwize} \end{cases} \tag{5.8}
$$

It follows from (5.8) and the Law of Total Probability that

$$
\begin{aligned}
&P(A \leq a,\ B \leq b) \\
&\quad = P(XY + Z \leq a,\ W \leq b,\ Y > 0) + P(Z \leq a,\ W - Y \leq b,\ Y \leq 0).
\end{aligned} \tag{5.9}
$$

The mixed nature of W and another application of the Law of Total Probability expands the righthand side of (5.9) into

$$
\begin{aligned}
&P(XY + Z \leq a,\ w^* = W \leq b,\ Y > 0) \\
&+ P(XY + Z \leq a,\ w^* < W \leq b,\ Y > 0) \\
&+ P(Z \leq a,\ w^* = W \leq b + Y,\ Y \leq 0) \\
&+ P(Z \leq a,\ w^* < W \leq b + Y,\ Y \leq 0).
\end{aligned} \tag{5.10}
$$

Taking advantage of X, Y and Z's continuity and W's limited continuity, (5.10) can be written with integrals and joint pdfs.

$$\begin{aligned} &P(w^* \le b) \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_0^{\infty} f_{W,XY,Y,Z}(w^*, \alpha, y, z) dy d\alpha dz \\ &+ \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_0^{\infty} \int_{w^*}^{b} f_{W,XY,Y,Z}(w, \alpha, y, z) dw dy d\alpha dz \\ &+ \int_{-\infty}^{a} \int_{w^*-b}^{0} f_{W,Y,Z}(w^*, y, z) dy dz \\ &+ \int_{-\infty}^{a} \int_{-\infty}^{0} \int_{w^*}^{b+y} f_{W,Y,Z}(w, y, z) dw dy dz. \end{aligned} \tag{5.11}$$

Transforming the integrals like in Section 5.1, (5.11) becomes

$$\begin{aligned} &P(w^* \le b) \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_0^{\infty} \frac{1}{y} f_{W,X,Y,Z}(w^*, \frac{\alpha}{y}, y, z) dy d\alpha dz \\ &+ \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_0^{\infty} \int_{w^*}^{b} \frac{1}{y} f_{W,X,Y,Z}(w, \frac{\alpha}{y}, y, z) dw dy d\alpha dz \\ &+ \int_{-\infty}^{a} \int_{w^*-b}^{0} f_{W,Y,Z}(w^*, y, z) dy dz \\ &+ \int_{-\infty}^{a} \int_{-\infty}^{0} \int_{w^*}^{b+y} f_{W,Y,Z}(w, y, z) dw dy dz. \end{aligned} \tag{5.12}$$

Accounting for independence in (5.12) gives the result for $F_{A,B}(a,b)$. Moreover, the continuity of X, Y and Z, the limited continuity of W and Leibniz's integral rule imply that $\frac{d^2}{dadb} F_{A,B}(a,b)$ makes sense for $b \ne w^*$. So (A,B) is mixed with continuous joint cdf except at $B = w^*$. It is also clear from (5.12) that $F_{A,B}(a,b) = 0$ for $b < w^*$. □

Theorem 9. *Let* $3 \le k \le n+1$*. Then* (TI_{k-1}, A_k) *is mixed and the only discontinuity* *in its cdf is at* $TI_{k-1} = T_1$*. Moreover,* $F_{TI_{k-1},A_k}(b,a) = 0$ *for* $b < T_1$ *and*

$$\begin{aligned} F_{TI_{k-1},A_k}(b,a) = P(T_1 \le b) &\int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_0^{\infty} \frac{1}{y} f_{TI_{k-2},A_{k-1}}(T_1,y) f_{X,Z}(\frac{\alpha}{y},z) dy d\alpha dz \\ &+ \int_{\mathbb{R}} \int_{-\infty}^{a-z} \int_0^{\infty} \int_{T_1}^{b} \frac{1}{y} f_{TI_{k-2},A_{k-1}}(w,y) f_{X,Z}(\frac{\alpha}{y},z) dw dy d\alpha dz \\ &+ F_Z(a) \int_{T_1-b}^{0} f_{TI_{k-2},A_{k-1}}(T_1,y) dy \\ &+ F_Z(a) \int_{-\infty}^{0} \int_{T_1}^{b+y} f_{TI_{k-2},A_{k-1}}(w,y) dw dy \end{aligned}$$

$$f_{TI_{k-1},A_k}(b,a) = \begin{cases} \frac{d^2}{dadb} F_{TI_{k-1},A_k}(b,a), & b \ne T_1 \\ \frac{d}{da} F_{TI_{k-1},A_k}(b,a), & b = T_1 \end{cases}$$

where

$$f_{X,Z}(x,z) = \frac{1}{|T_{k-1}|} f_{X_{\cdot 1},X_{\cdot 2}}(\frac{z+T_k}{T_{k-1}}, x).$$

Proof. Set $W = TI_{k-2}$, $X = X_{k-1,2}$, $Y = A_{k-1}$ and $Z = X_{k-1,1}T_{k-1} - T_k$. From the recursive definitions for value averaging, Lemma 4's dependency requirements are satisfied. X and Z are continuous by their definitions. Y is continuous by Theorem 8.

When $k = 3$, $W = T_1$, so (W,Y) is mixed, with the only discontinuity at $W = T_1$ and $F_{W,Y}(w,y) = 0$ for $w < T_1$ (this is necessary to start the induction). An application of Lemma 4 finishes the induction, showing that (TI_{k-1}, A_k) is mixed and the only discontinuity in its cdf is at $TI_{k-1} = T_1$; moreover, $F_{TI_{k-1},A_k}(b,a) = 0$ for $b < T_1$. Applying Lemma 4 with the recursive definitions of A_k and TI_{k-1} yields the result. The pdf $f_{X,Y}$ was derived in Section 5.1. □

5.3 $VARR_n$ Cumulative distribution function

Theorem 10. *The cdf of* $VARR_n$ *is given by*

$$\begin{aligned} P(VARR_n \le v) = &\int_0^{\infty} \int_{-\infty}^{v} b f_{TI_n,A_{n+1}}(b,(\alpha+1)b - T_{n+1}) d\alpha db \\ &+ \int_{-\infty}^{v} T_1 f_{TI_n,A_{n+1}}(T_1,(\alpha+1)T_1 - T_{n+1}) d\alpha. \end{aligned}$$

Proof. Using the definition of A_n to substitute for $VoBI_n + VoSI_n$,

$$VARR_n = \frac{A_{n+1} + T_{n+1}}{TI_n} - 1.$$

So the cdf of $VARR_n$ is found by bootstrapping off the cdf of (TI_n, A_{n+1}), which is obtained through recursion and Theorem 9.

From the Law of Total Probability,

$$\begin{aligned} P(VARR_n \leq v) = P(&\frac{A_{n+1} + T_{n+1}}{TI_n} - 1 \leq v,\ TI_n > T_1) \\ &+ P(\frac{A_{n+1} + T_{n+1}}{TI_n} - 1 \leq v,\ TI_n = T_1). \end{aligned} \tag{5.13}$$

Accounting for the mixed nature of (TI_n, A_{n+1}), (5.13) is rewritten with integrals and pdfs.

$$\begin{aligned} P(VARR_n \leq v) = &\int_0^\infty \int_{-\infty}^v f_{TI_n, \frac{A_{n+1}+T_{n+1}}{TI_n} - 1}(b, \alpha) d\alpha db \\ &+ \int_{-\infty}^v f_{TI_n, \frac{A_{n+1}+T_{n+1}}{TI_n} - 1}(T_1, \alpha) d\alpha. \end{aligned} \tag{5.14}$$

Transforming the integral in (5.14) like in Section 4.1 gives the result. □

5.4 TI_k Expectation and variance

The expectation of TI_k can be expressed recursively using Theorem 8. By definition, $TI_k = TI_{k-1} + g(-A_k)$. It follows from the linearity and definition of expectation that

$$\begin{aligned} E[TI_k] &= E[TI_{k-1}] + E[g(-A_k)] \\ &= E[TI_{k-1}] - \int_{-\infty}^0 a f_{A_k}(a) da. \end{aligned}$$

Alternatively, the expectation and variance of TI_k can be expressed directly using Theorem 9.

$$\begin{aligned} E[TI_k] &= \int_{T_1}^\infty b \int_{\mathbb{R}} f_{TI_{k-1}, A_k}(b, a) da db \\ Var(TI_k) &= \int_{T_1}^\infty b^2 \int_{\mathbb{R}} f_{TI_{k-1}, A_k}(b, a) da db - E[TI_{k-1}]^2. \end{aligned}$$

5.5 $VARR_n$ Expectation and variance

The expectation and variance of $VARR_n$ can be expressed directly using Theorem 8.

$$\begin{aligned} E[VARR_n] &= \int_{\mathbb{R}} v f_{VARR_n}(v) dv \\ Var(VARR_n) &= \int_{\mathbb{R}} v^2 f_{VARR_n}(v) dv - E[VARR_n]^2. \end{aligned}$$

6. DISTRIBUTION OF REAL RETURNS

S&P and 1YTR real returns are tested separately for autoregression of order 2. Then residuals are fit to a bivariate random variable. The Normality of residuals is assessed. The 95% confidence standard is used to make decisions about autoregression and Normality of residuals. An alternative kernel density estimate for residuals is developed. Results under the kernel density estimate and Normality assumption will be compared in applications.

6.1 Autoregression of real returns

The Ljung-Box test with lag 2 produces p-values greater than 0.2 for pre- and post-1951 S&P, so the null hypothesis that the data is AR(0) is not rejected. The same test produces p-values less than 0.007 for pre- and post-1951 S&P, so the null hypothesis that the data is AR(0) is rejected. Therefore S&P real returns will be modeled as AR(0) and 1YTR real returns will be modeled as AR(1) or AR(2).

Fitting pre- and post-1951 1YTR to AR(2) produces insignificant p-values (> 0.3) for the coefficient of the second lag. As a result, AR(2) is rejected. Fitting AR(1) yields significant p-values (< 0.009) for the coefficient of the first lag, so AR(1) is selected as the model for 1YTR real returns. Results are given in (6.1).

$$\begin{aligned} \text{pre-1951 S\&P RR:} \quad & s_k = 0.081 + \epsilon_k^{s1} \\ \text{post-1951 S\&P RR:} \quad & s_k = 0.085 + \epsilon_k^{s2} \\ \text{pre-1951 1YTR RR:} \quad & b_k = 0.024 + 0.296 b_{k-1} + \epsilon_k^{b1} \\ \text{post-1951 1YTR RR:} \quad & b_k = 0.005 + 0.663 b_{k-1} + \epsilon_k^{b2} \end{aligned} \tag{6.1}$$

Before 1951, the mean and covariance for the bivariate random variable $[\epsilon_k^{s1}, \epsilon_k^{b1}]^T$

are

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.0354 & 0.0010 \\ 0.0010 & 0.0065 \end{bmatrix}. \tag{6.2}$$

After 1951, the mean and covariance for the bivariate random variable $[\epsilon_k^{s2}, \epsilon_k^{b2}]^T$ are

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.0265 & 0.0003 \\ 0.0003 & 0.0004 \end{bmatrix}. \tag{6.3}$$

6.1.1 Normality of residuals

The residuals of each model in (6.1) produces a p-value greater than 0.3 from the Kolmogorov-Smirnov test. This indicates that residuals can be modeled with a Normal random variable. Quantile-quantile plots are presented to further asses Normality and decide whether a Normal assumption is justified. Figures 6.1 and 6.2 show that S&P and 1YTR residuals align well with Normality, but 1YTR has more deviations. Overall, the deviations from Normality are minor, and the QQ plots support the Kolmogorov-Smirnov tests. A Normal assumption of residuals is justified.

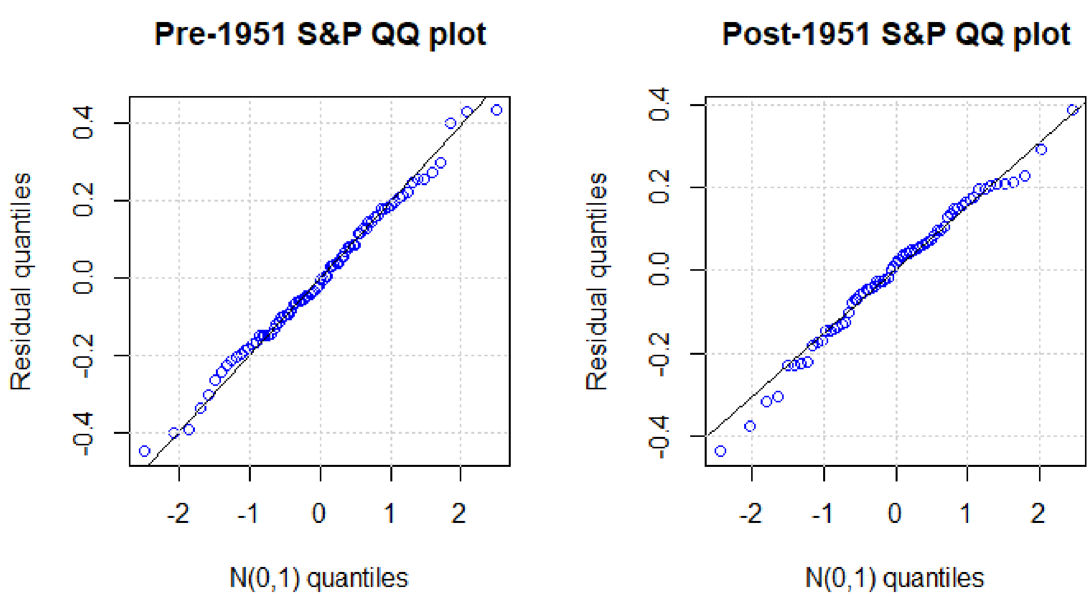

Fig. 6.1: Quantile-quantile plots of S&P residuals (see (6.1)) vs the standard Normal distribution in two sets of years: 1871-1950 and 1951-2019. The black line indicates where quantile coordinates should be if the unknown distribution is Normal.

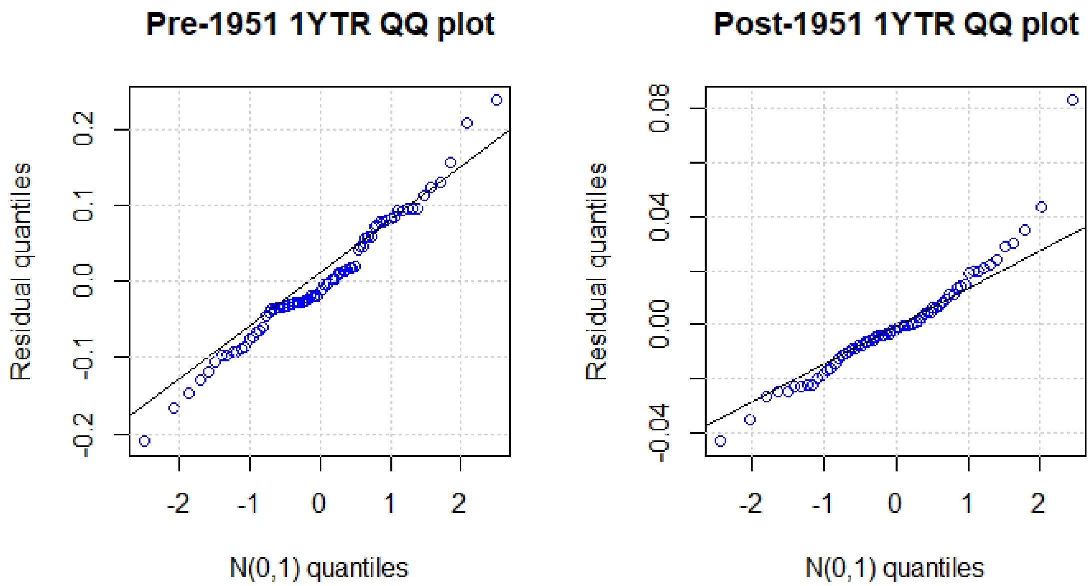

Fig. 6.2: Quantile-quantile plots of 1YTR residuals (see (6.1)) vs the standard Normal distribution in two sets of years: 1871-1950 and 1951-2019. The black line indicates where quantile coordinates should be if the unknown distribution is Normal.

6.1.2 Bivariate kernal density estimation of residuals

While residuals do align well with Normality, there are still deviations, and it is easy to construct a kernel density estimate for comparison. Denote the bivariate Gaussian kernel with $K(x) = (2\pi)^{-1}\exp(-\frac{1}{2}x^T x)$. Then the density estimate is given by the following equation.

$$\hat{f}(x) = \frac{1}{n\sqrt{|H|}} \sum_{i=1}^{n} K(H^{-\frac{1}{2}}(x - \begin{bmatrix} X_{i1} \\ X_i 2 \end{bmatrix})). \tag{6.4}$$

H is the plug-in bandwidth, defined as $\arg\min_{H \in \mathcal{H}} \text{AMISE}\hat{f}(\cdot; H)$, where $\mathcal{H}$ is the space of symmetric, positive-definite 2×2 matrices and

$$\text{AMISE}\hat{f}(\cdot; H) = \frac{R(K)}{n\sqrt{|H|}} + \frac{m_2(K)^2}{4}(\text{vech}^T H)\Psi_4(\text{vech} H). \tag{6.5}$$

In (6.5),

$$
\begin{aligned}
R(K) &= \int_{\mathbb{R}^2} K(x)^2 dx \\
m_2(K)I_2 &= \int_{\mathbb{R}^2} xx^T K(x) dx \\
\text{vech} H &= [h_{11}\ h_{12}\ h_{22}] \\
\Psi_4 &= \int_{\mathbb{R}^2} \text{vech}[2D^2 f(x) - \text{dg} D^2 f(x)] \text{vech}^T [2D^2 f(x) - \text{dg} D^2 f(x)] dx
\end{aligned}
\tag{6.6}
$$

where $D^2 f(x)$ denotes the Hessian matrix and dg is an operator performed on a matrix to set non-diagonal elements to zero.

H is selected with the goal of minimizing MISE$\hat{f}(\cdot; H)$. AMISE is used instead of MISE because it approximates MISE and is more wieldy [9].

Since the K is the Gaussian kernel here, $R(K) = \frac{1}{4\pi}$ and $m_2(K) = 1$ (see Appendix). Hence, (6.5) reduces to

$$
\text{AMISE}\hat{f}(\cdot; H) = \frac{1}{4\pi n \sqrt{|H|}} + \frac{1}{4}(\text{vech}^T H)\Psi_4(\text{vech} H). \tag{6.7}
$$

Bandwidth matrices for the bivariate random variable

$$
[\text{S\&P residuals},\ \text{1YTR residuals}]^T
$$

are given below. H_1 is for the pre-1951 density, and H_2 is for the post-1951 density.

$$
\begin{aligned}
H_1 &= \begin{bmatrix} 0.0087623046 & 0.0003308656 \\ 0.0003308656 & 0.0012262923 \end{bmatrix} \\
H_2 &= \begin{bmatrix} 0.0072875595 & 0.0001173202 \\ 0.0001173202 & 0.00007379545 \end{bmatrix}
\end{aligned}
\tag{6.8}
$$

Figure 6.3 shows the density estimates for residuals.

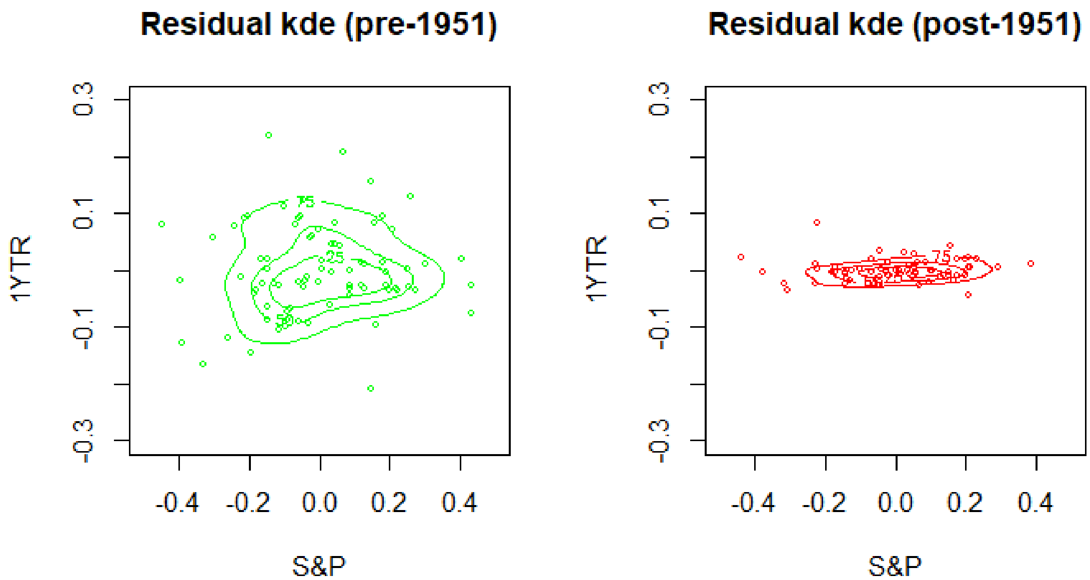

Fig. 6.3: Bivariate kernel density estimate of S&P and 1YTR residuals as described by (6.1) and (6.4). The 25%, 50% and 75% quantiles are outlined. For comparison, the actual bivariate residuals are plotted as points.

7. APPLICATIONS

The simplicity of dollar-cost averaging led to more applicable theoretical results. Specifically, the expectation, standard deviation, Sharpe ratio and related optimization are easily computed. None of the theoretical results on recursively defined distributions are applied because they take too much time and space to compute with accuracy. The value averaging theoretical results consist entirely of recursively defined distributions, so there are no applications there. In order to make statements about their distributions, the dollar-cost and value averaging strategies are simulated. From there, samples are used to estimate quantiles, expectation, standard deviation, Sharpe ratio and related optimization. The code used to produce figures is on github [5].

7.1 Applications of DCA theoretical results

Dollar-cost averaging theoretical results require AR(0) real returns, which implies real returns are iid with constant mean and covariance. The following applications make use of the sample mean and covariance for the pre-1951 and post-1951 time periods, given in (7.1) and (7.2). Recall that p indicates the proportion invested in S&P each year.

Before 1951, the mean and covariance for the bivariate random variable $[s_k + 1, b_k + 1]^T$ are

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 1.0808 \\ 1.0344 \end{bmatrix}, \quad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.0350 & 0.0018 \\ 0.0018 & 0.0070 \end{bmatrix}. \tag{7.1}$$

After 1951, the mean and covariance for the bivariate random variable $[s_k+1, b_k+1]^T$ are

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 1.0850 \\ 1.0147 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.0262 & 0.0005 \\ 0.0005 & 0.0007 \end{bmatrix}. \tag{7.2}$$

7.1.1 *Expectation*

Figure 7.1 shows the expectation of $DCRR_n$ for a range of n. Observe the linearity of expectation. The value p at which the post-1951 expectation overtakes the pre-1951 expectation decreases as the investment length increases.

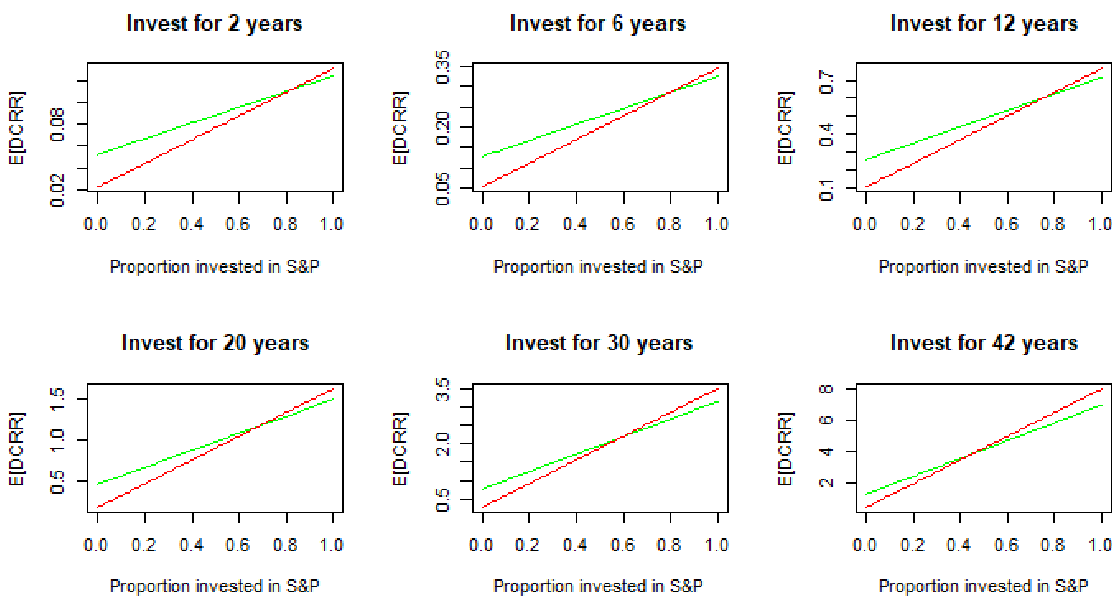

Fig. 7.1: Computed using Theorem 2 and the sample mean and covariance for two time periods. Green indicates pre-1951 expectations. Red indicates post-1951 expectations.

7.1.2 *Standard deviation*

Figure 7.2 shows the standard deviation of $DCRR_n$ for a range of n. Regardless of investment length, the standard deviation is higher with the pre-1951 period. Post-1951, there is a linear trend in standard deviation. Pre-1951, there is increased convexity for small p, but a linear trend develops as p increases. That stretch of convexity is less and less noticeable as the length of investment increases.

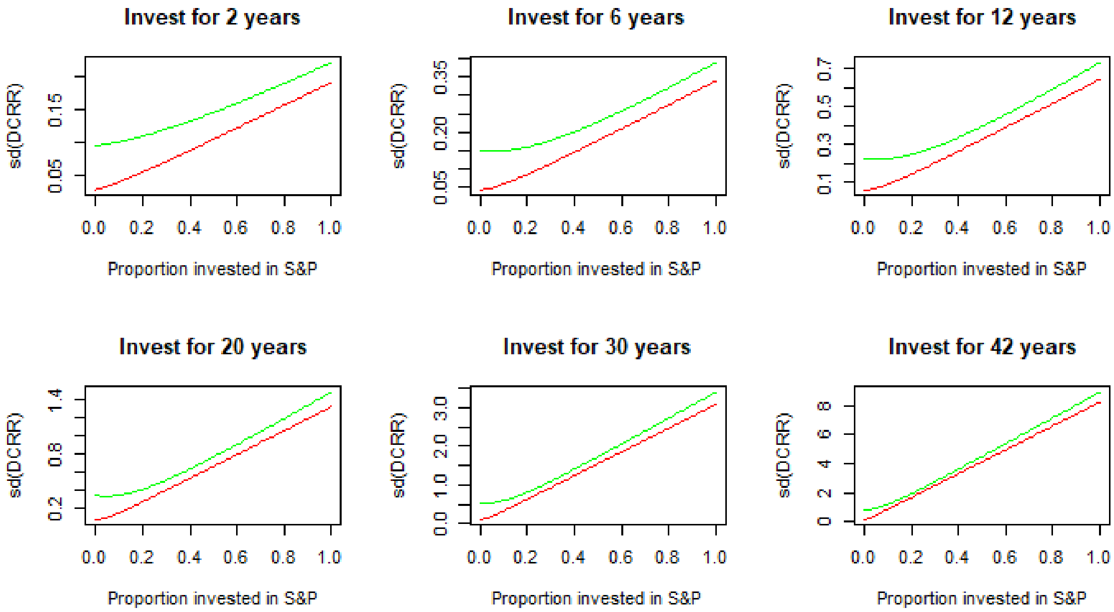

Fig. 7.2: Computed using Theorem 2 and the sample mean and covariance for two time periods. Green indicates pre-1951 standard deviations. Red indicates post-1951 standard deviations.

Figure 7.3 shows $\arg\min_p Var(DCRR_n)$ - which is the same with standard deviation - for a range of n. All investment lengths have small argmins, close to 0. Pre-1951, with investment length greater than 4 years, it is less volatile to invest a small proportion in S&P than to invest everything in 1YTR. However, Figure 7.2 shows that this difference in standard deviation was small.

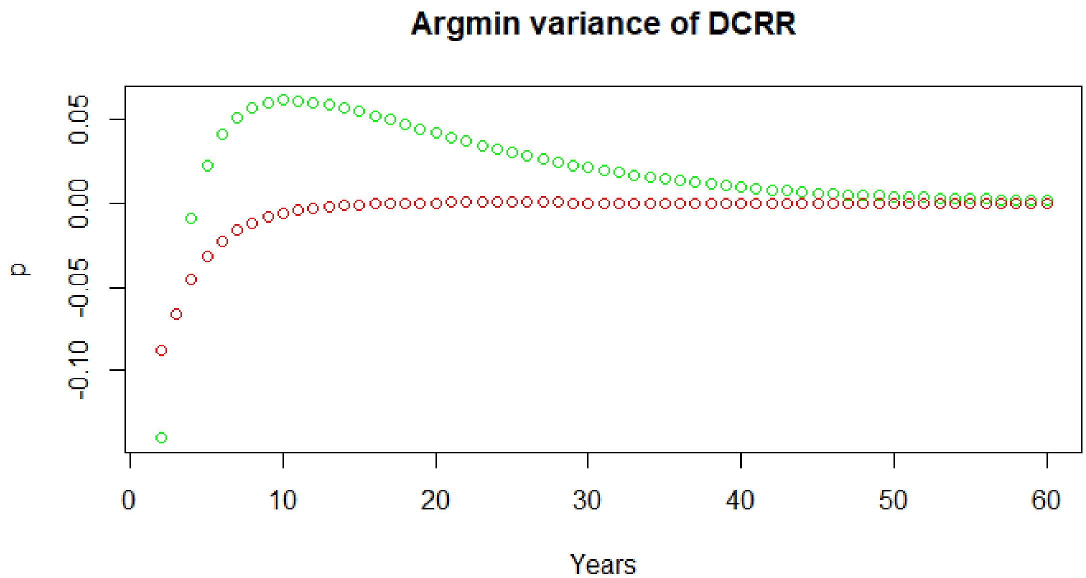

Fig. 7.3: Computed using Theorem 6, Theorem 2 and the sample mean and covariance for two time periods. Green indicates pre-1951 argmins. Red indicates post-1951 argmins.

7.1.3 Sharpe ratio

Figure 7.4 shows the Sharpe ratio of $DCRR_n$ for a range of n. The post-1951 period has a higher Sharpe ratio for investment lengths of 2, 6 and 12 years. Starting with an investment length of 20 years, the pre-1951 period begins to have a higher Sharpe ratio at some p. The range of p at which the Sharpe ratio of the pre-1951 period is higher increases as the investment length increases past 20 years. Regardless of time period, the Sharpe ratio tends to level out as the proportion invested in S&P increases.

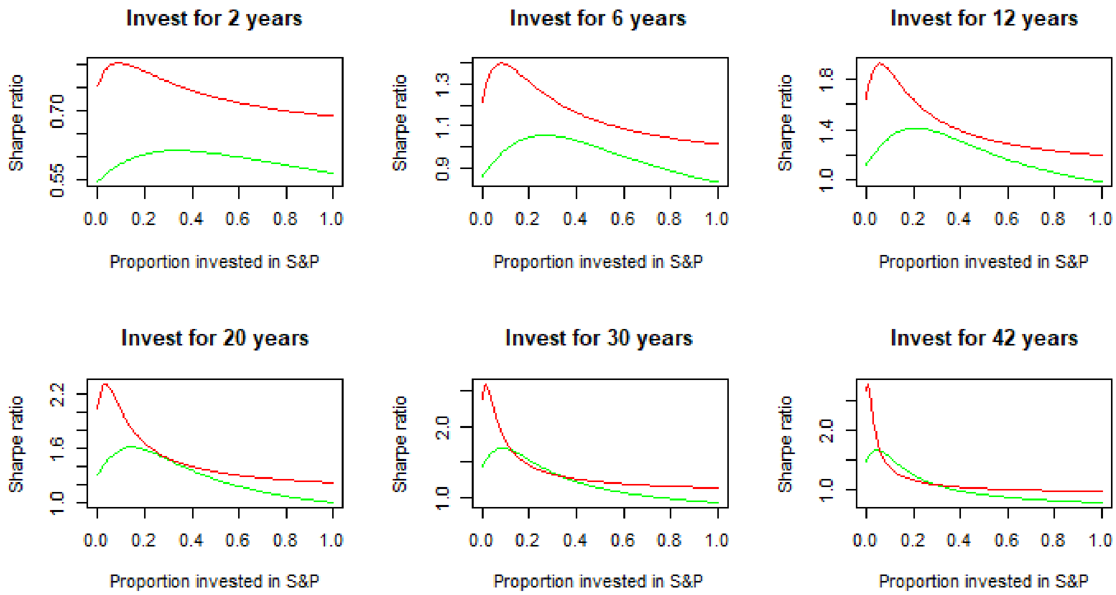

Fig. 7.4: Computed using Theorem 2 and the sample mean and covariance for two time periods. Green indicates pre-1951 Sharpe ratios. Red indicates post-1951 Sharpe ratios.

Figure 7.5 shows the Sharpe ratio argmax of $DCRR_n$ for a range of n. The difference in argmax between the two time periods decreases as the investment length increases. Pre-1951 has a larger argmax for each investment length. Regardless of investment length or time period, the argmax occurs at $p < 0.4$. Furthermore, the argmax decreases close to 0 as investment length increases.

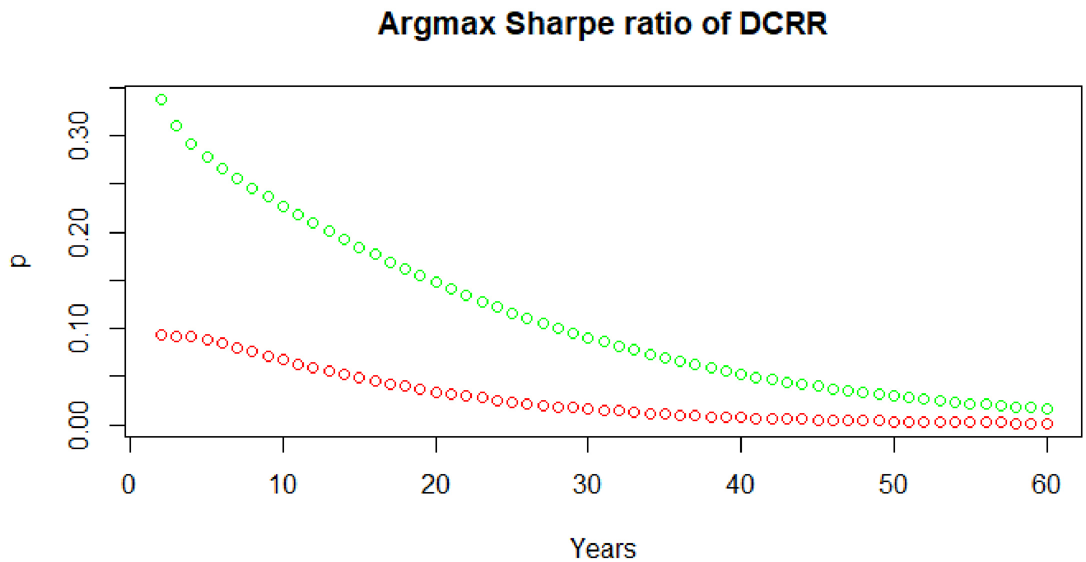

Fig. 7.5: Computed using Theorem 7, Theorem 2 and the sample mean and covariance for two time periods. Green indicates pre-1951 argmaxes. Red indicates post-1951 argmaxes. p indicates the proportion invested in S&P.

Figure 7.6 shows the Sharpe ratio max of $DCRR_n$ for a range of n. The post-1951 Sharpe ratio max continues to grow, whereas the pre-1951 max levels out as the length of investment increases. In general, the post-1951 time period offers higher Sharpe ratios.

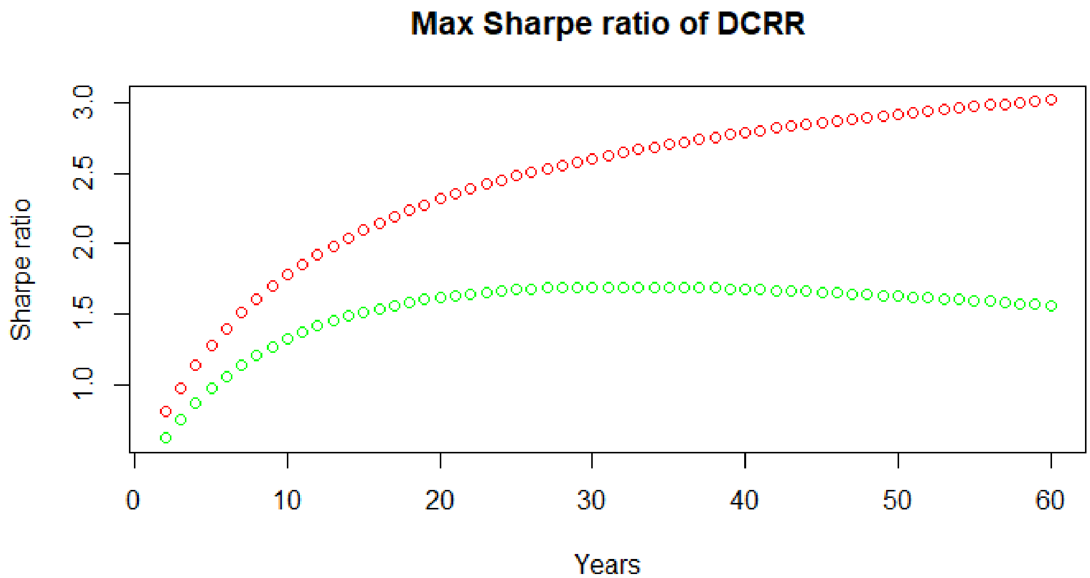

Fig. 7.6: Computed using Theorem 7, Theorem 2 and the sample mean and covariance for two time periods. Green indicates pre-1951 maxes. Red indicates post-1951 maxes.

7.1.4 Summary

There is a proportion invested in S&P where both time periods produce the same expectation. Moreover, post-1951 expectations are higher than pre-1951 expectations for larger proportions invested in S&P. The opposite is true for smaller proportions. When the proportion is fixed, expectation is sensative to time period and investment length. Longer investments and larger proportions invested in S&P lead to greater expectations and standard deviations. In other words, return and volatility increases with investment length and the proportion invested in S&P.

Longer investments offer higher Sharpe ratios, implying that risk-return tradeoff improves as investment length increases. Investing a small proportion in S&P offers better risk-return tradeoff than investing everything into 1YTR. Alternatively, the risk-return tradeoff gained by investing a small proportion into 1YTR decreases as investment length increases. For longer investments, it can be better to focus exclusively on S&P.

7.2 *Dollar-cost averaging simulation*

The dollar-cost averaging strategy is simulated using the bivariate random variables fitted in Chapter 6. The time-span is split into two sets, 1871-1950 and 1951-2019, in an attempt to account for the apparent change in 1YTR RR variance starting at 1951. 0.975, 0.5 and 0.025 quantiles for real return are plotted in Figures 7.7, 7.8 and 7.9. The kernel density estimate of residuals has a tendency to create a slightly larger 0.975 quantile and smaller 0.025 quantile compared to the Normal residuals. So kernel density estimates of residuals produce larger confidence intervals. However, the difference is not egregious. Furthermore, the shape of each quantile (as a function of proportion) is very similar regardless of how residuals were generated.

Figure 7.9 indicates that it is least risky when the proportion invested in S&P is between 0.1 and 0.5, depending on investment length. This reference to risk is different from the Sharpe ratio in that it minimizes probability of incurring a loss and does not factor in possible gains. Figures 7.7 and 7.8 show that the 0.975 and 0.5 quantiles are approximately linear over p. Similar to how the kernel density estimate produced larger confidence intervals, pre-1951 confidence intervals are larger than post-1951 confidence intervals.

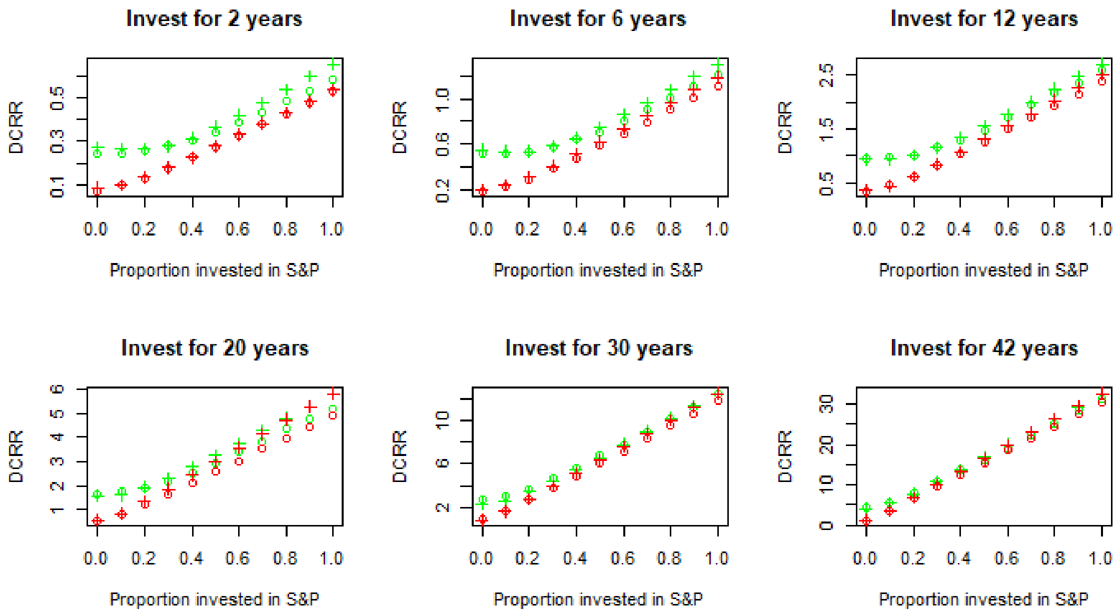

Fig. 7.7: DCA upper bound of 95% confidence interval for total real returns after investing for 2, 6, 12, 20, 30 and 42 years using the dollar-cost averaging strategy. In each plot, the proportion p of the fixed annual investment placed in S&P is varied from 0 to 1 in 0.1 size intervals; the remaining proportion is invested in 1YTR. Green indicates the residuals for S&P and 1YTR used in computation were simulated using a (pre-1951) bivariate random variable. Red indicates the residuals were simulated using a (post-1950) bivariate random variable. A circle indicates the bivariate random variable was Normal, described by (6.2). A plus indicates the bivariate random variable was estimated by (6.4) and (6.8). 3000 simulations were conducted for each time period and proportion invested in S&P.

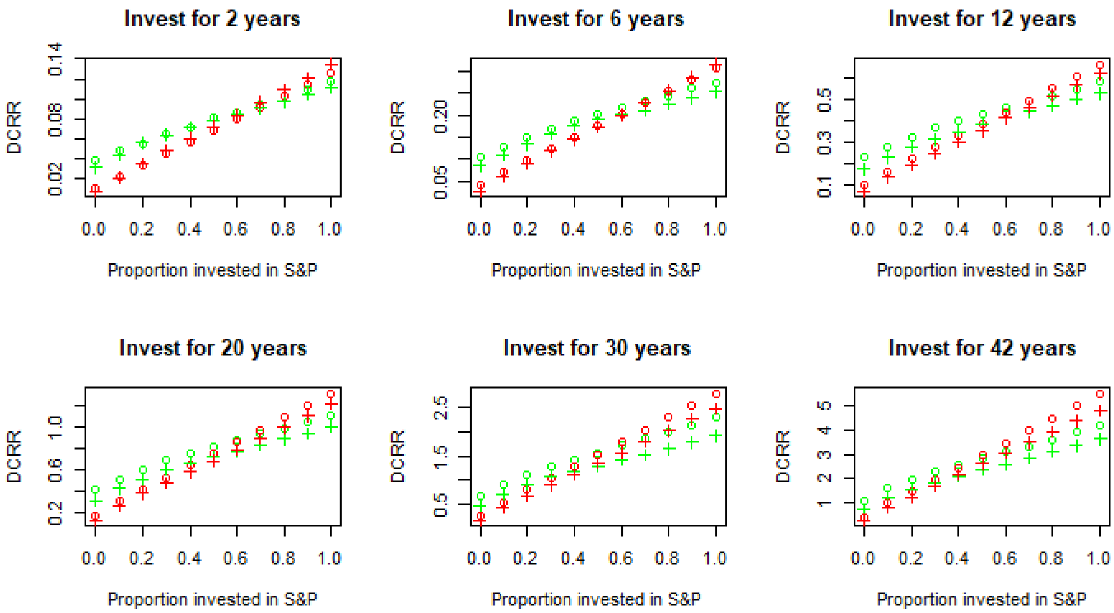

Fig. 7.8: DCA 0.5 quantile with the same details as in Figure 7.7.

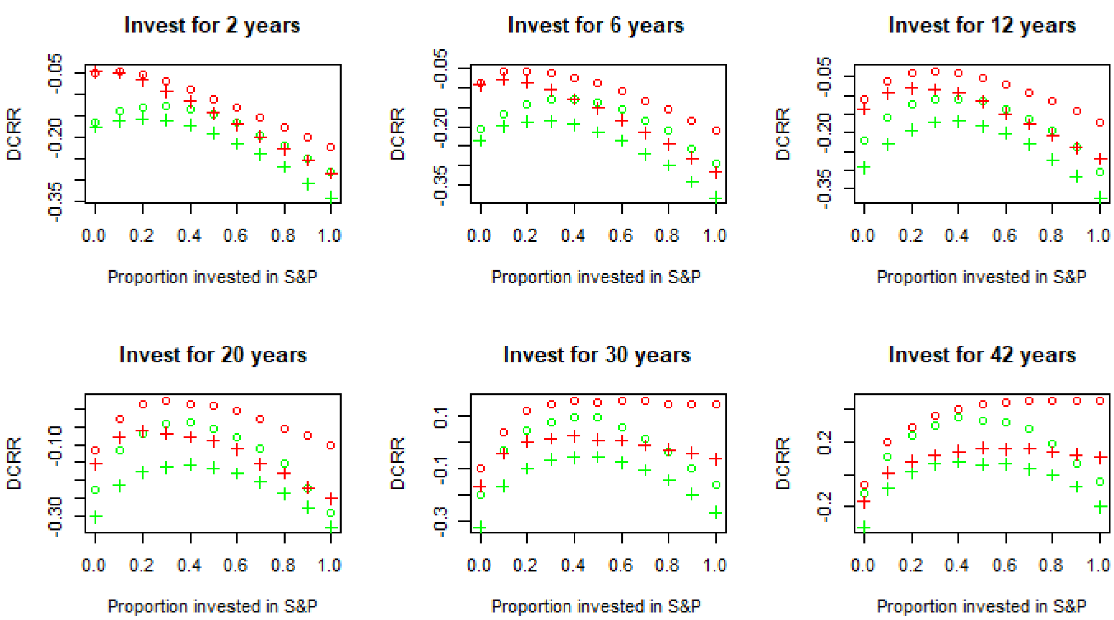

Fig. 7.9: DCA lower bound of 95% confidence interval with the same details as in Figure 7.7.

Expectation, standard deviation and Sharpe ratio are plotted in Figures 7.10, 7.11 and 7.12. The difference resulting from generating residuals via kernel density estimate or Normal density is not egregious. Furthermore, shape is nearly identical for each investment length.

The expectations in Figure 7.10 tend to be larger than the 0.5 quantiles in Figure 7.8, indicating right skew; the difference is most significant at higher rates. Theoretical expectations and standard deviations are almost identical to their simulated versions, which account for autoregression. Theoretical Sharpe ratios tend to exceed their simulated versions, but the shape and argmax is close. The Normal assumption tends to exaggerate Sharpe ratios compared with the kernel density estimate. In addition, the argmax of the 0.025 quantile tends to be larger than the argmax Sharpe ratio.

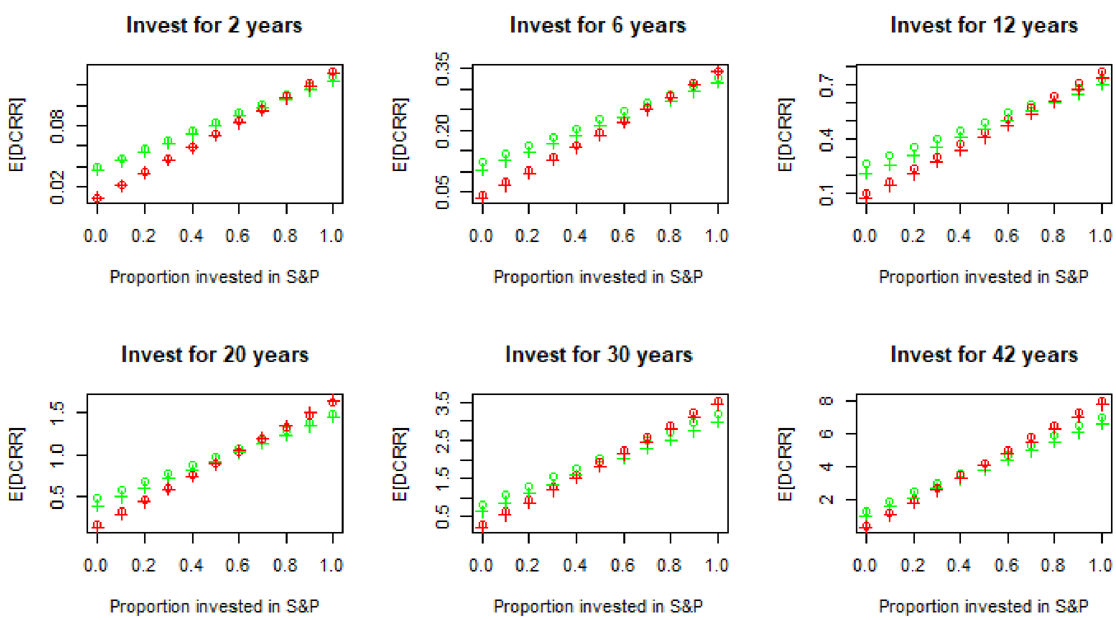

Fig. 7.10: DCA expectation with the same details as in Figure 7.7.

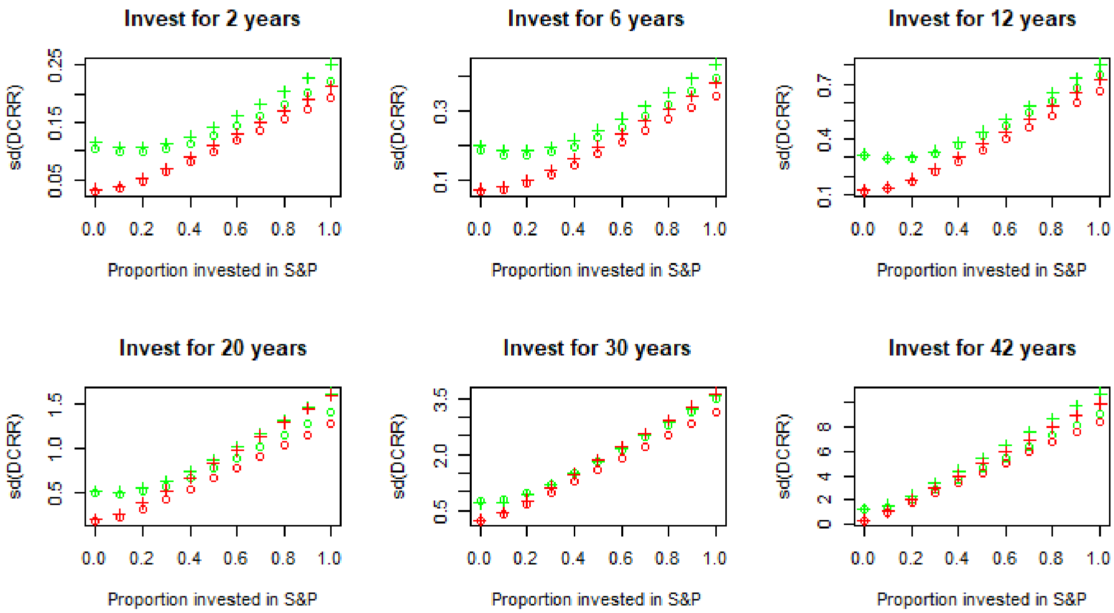

Fig. 7.11: DCA standard deviation with the same details as in Figure 7.7.

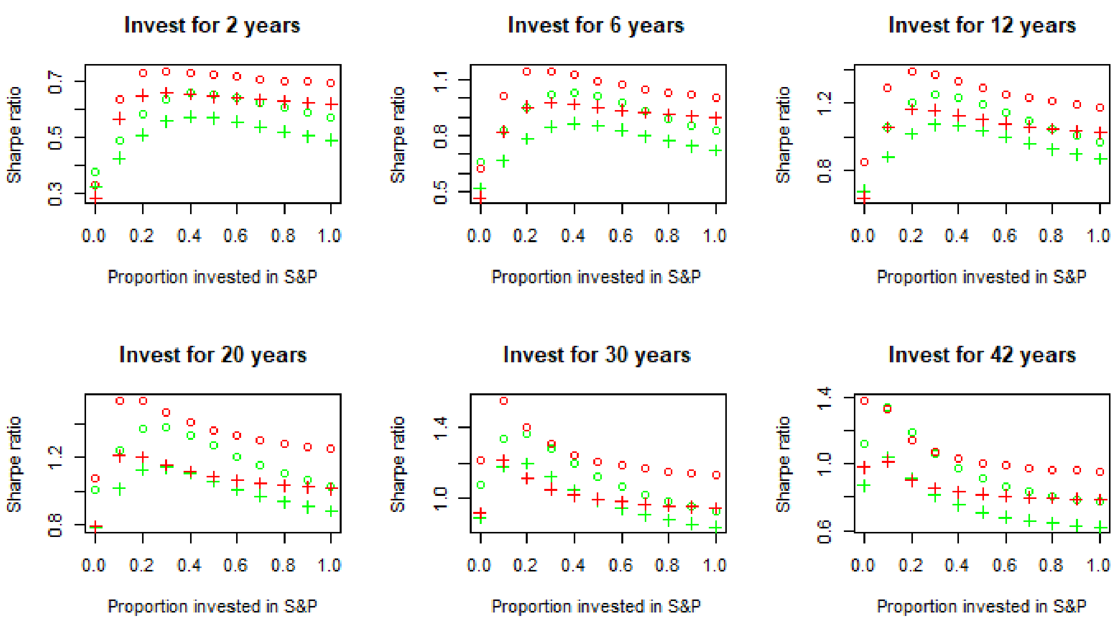

Fig. 7.12: DCA Sharpe ratio with the same details as in Figure 7.7.

7.2.1 Summary

Pre-1951 confidence intervals are slightly larger than post-1951 confidence intervals. Using Normal residuals produces larger confidence intervals than kernel density estimates, but the difference is not eggregious. When quantiles are graphed over the proportion, shape is about the same for Normal and kernel density estimates. Overall, the Normal assumption of residuals produces similar results.

Regardless of investment length or proportion invested in S&P, real returns are right skewed. If the goal is to minimize probability of incurring a loss, the optimal proportion invested in S&P may be larger than the argmax Sharpe ratio. When the proportion is fixed, quantiles increase with investment length. So longer investments are less risky and offer better returns.

Assuming real returns are AR(0) despite evidence to the contrary does not have a major impact on expectation, standard deviation and Sharpe ratio. Furthermore, investing decisions can be made using the AR(0) assumption, as the decisions are similar after accounting for autoregression of higher order.

7.3 Value averaging simulation

The value averaging strategy is simulated using the bivariate random variables fitted in Chapter 6. The time-span is split into two sets, 1871-1950 and 1951-2019, in an attempt to account for the apparent change in 1YTR RR variance starting at 1951. 0.975, 0.5 and 0.025 quantiles for real return are plotted in Figures 7.13, 7.14 and 7.15. The kernel density estimate of residuals has a tendency to create a slightly larger 0.975 quantile and smaller 0.025 quantile compared to the Normal residuals. So kernel density estimates of residuals produce larger confidence intervals. However, the difference is not egregious. Furthermore, the shape of each quantile (as a function of rate) is very similar regardless of how residuals were generated.

Figure 7.15 indicates that it is least risky, in terms of incurring a loss, when the rate is small and investment length is over 20 years. For shorter investments, the 0.025 quantile is approximately constant over r, so it is best to choose a rate with high potential return. Figures 7.13 and 7.15 show that the 0.975 and 0.025 quantiles are approximately linear over r until the investment length passes 20 years. When the investment length is greater than 6 years, there is a value at which potential returns worsen as the rate increases past that value. Similar to how the kernel density estimate produced larger confidence intervals, pre-1951 confidence intervals are larger than post-1951 confidence intervals.

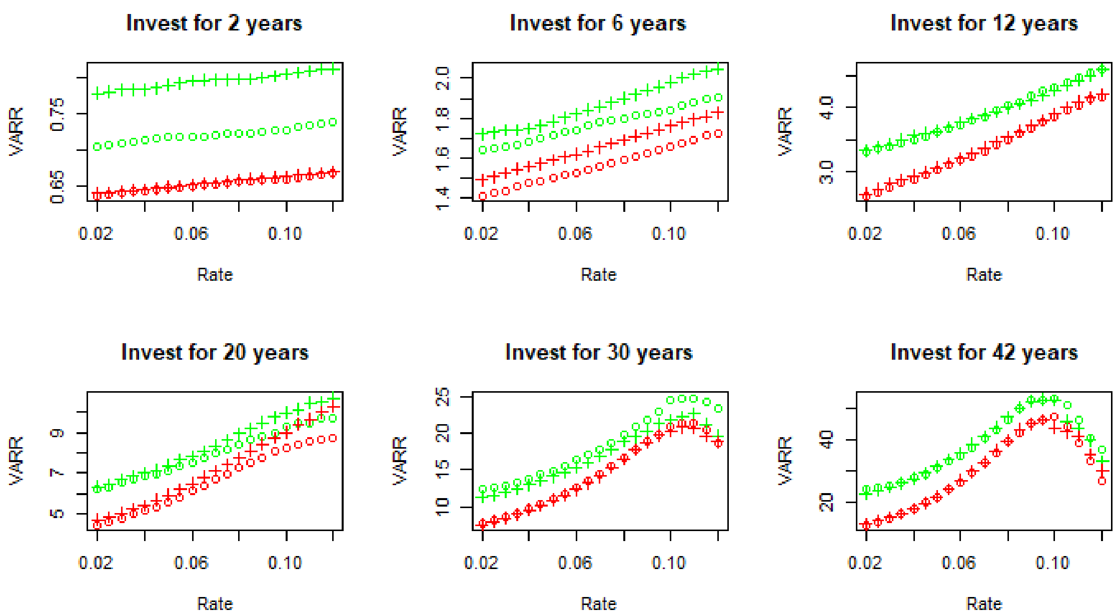

Fig. 7.13: VA upper bound of 95% confidence interval for total real returns after investing for 2, 6, 12, 20, 30 and 42 years using the value averaging strategy. In each plot, the rate r is varied from 0.02 to 0.12 in 0.005 size intervals. Green indicates the residuals for S&P and 1YTR used in computation were simulated using a (pre-1951) bivariate random variable. Red indicates the residuals were simulated using a (post-1950) bivariate random variable. A circle indicates the bivariate random variable was Normal, described by (6.2). A plus indicates the bivariate random variable was estimated by (6.4) and (6.8). 3000 simulations were conducted for each time period and rate.

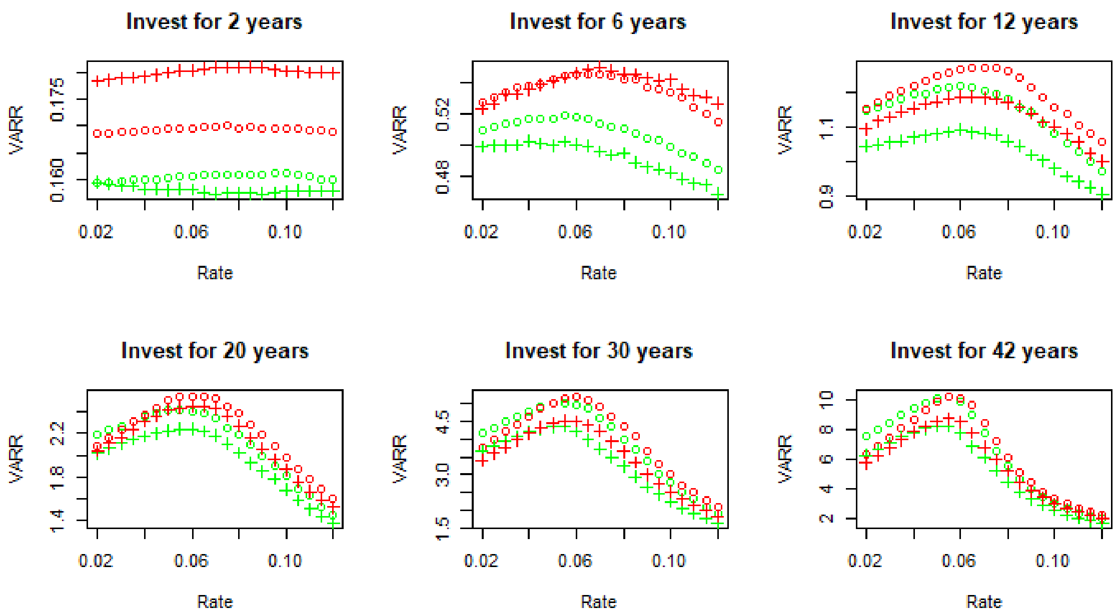

Fig. 7.14: VA 0.5 quantile with the same details as in Figure 7.13.

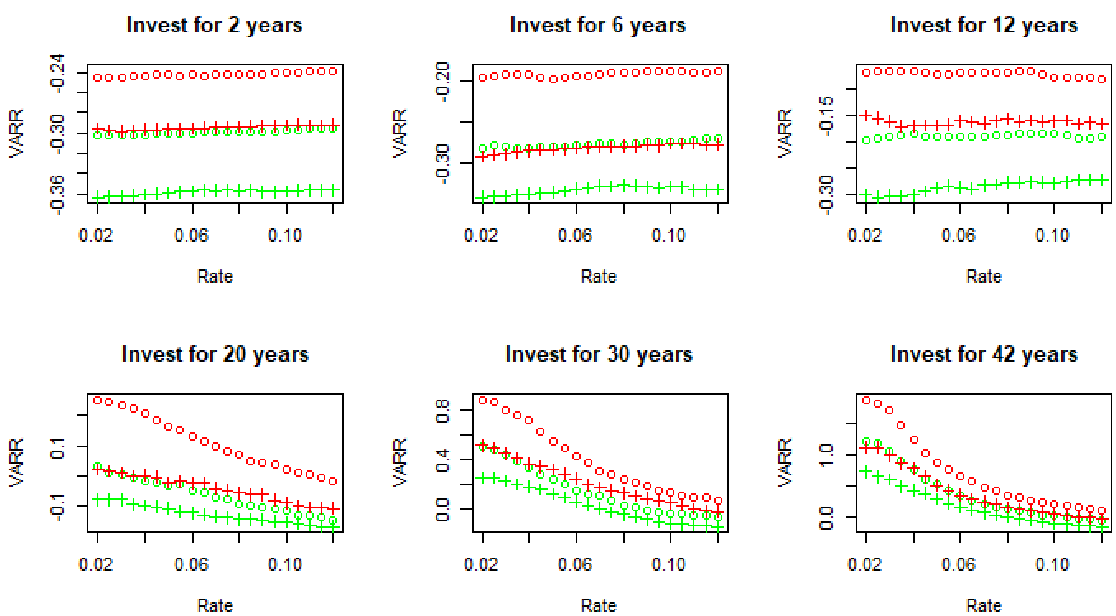

Fig. 7.15: VA lower bound of 95% confidence interval with the same details as in Figure 7.13.

Expectation, standard deviation and Sharpe ratio are plotted in Figures 7.16, 7.17 and 7.18. The difference resulting from generating residuals via kernel density estimate or Normal density is not egregious. Furthermore, shape is nearly identical for each investment length.

The expectations in Figure 7.16 tend to be larger than the 0.5 quantiles in Figure 7.14, indicating right skew. The difference is most significant at higher rates. In terms of the least risky investment being at small rates, the Sharpe ratios in Figure 7.18 agree with the 0.025 quantiles in Figure 7.15. The post-1951 Sharpe ratios are higher, suggesting that the risk-return tradeoff improved after 1951. Pre-1951 standard deviations are higher, which explains the larger confidence intervals.

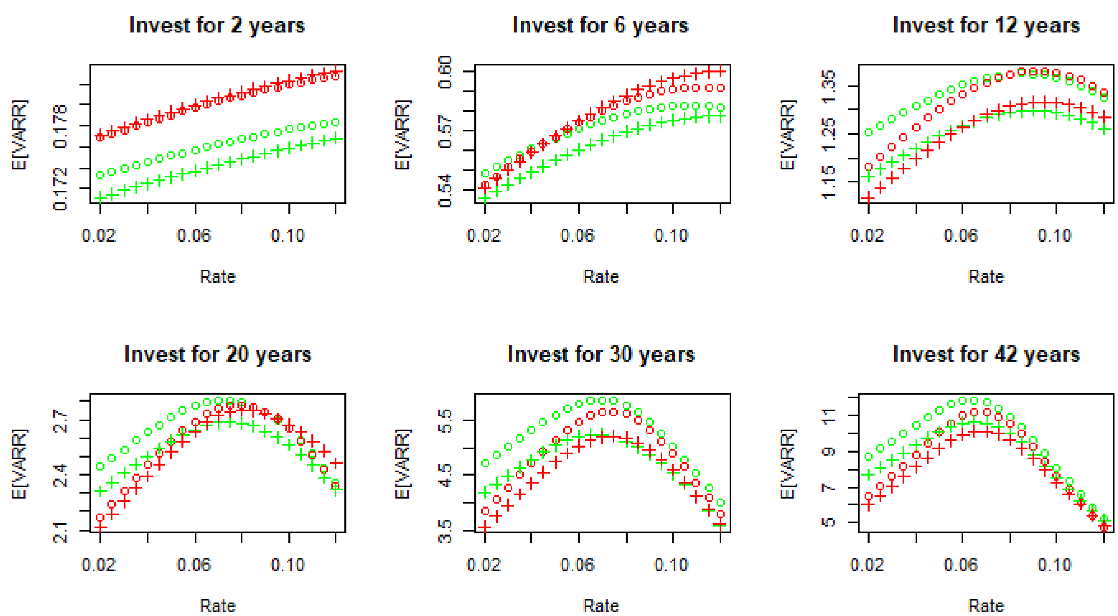

Fig. 7.16: VA expectation with the same details as in Figure 7.13.

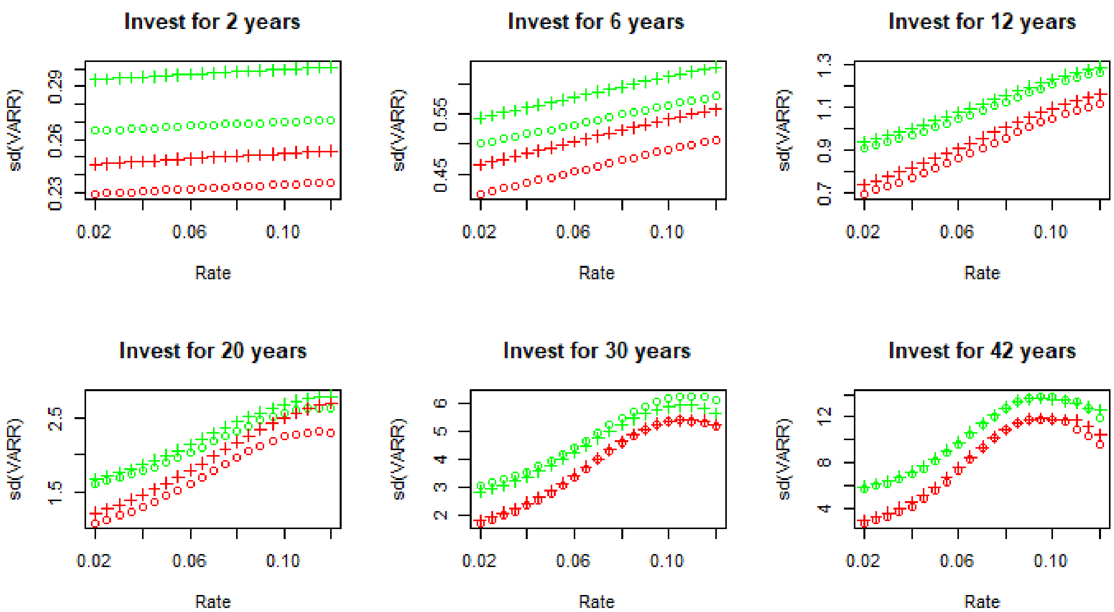

Fig. 7.17: VA standard deviation with the same details as in Figure 7.13.

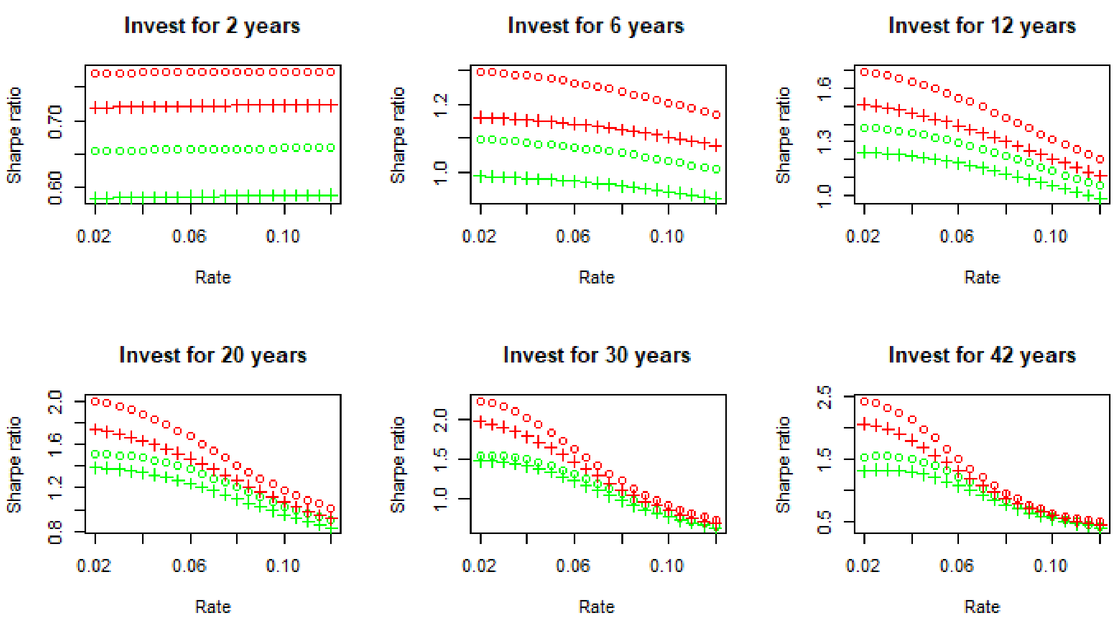

Fig. 7.18: VA Sharpe ratio with the same details as in Figure 7.13.

Figure 7.19 shows 0.975, 0.5 and 0.025 quantiles of the annual investment, placed at the beginning of the year. For a given year, the annual investment is not pulling from the bottomless cash reserve most of the time.

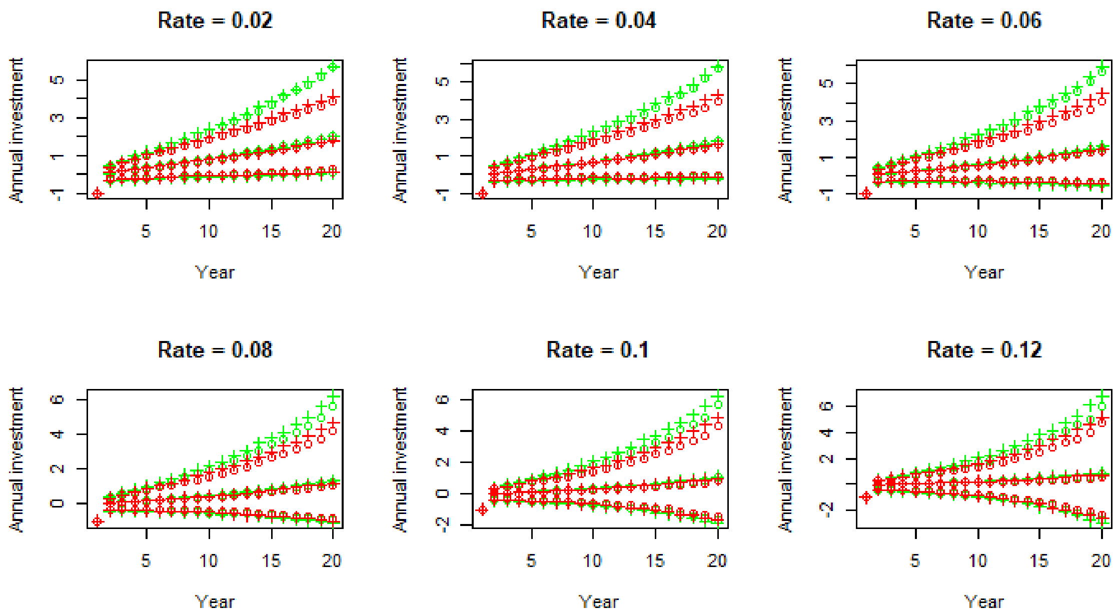

Fig. 7.19: A_k is plotted for six different rates, where k is the year. Positive A_k indicates the annual investment does not pull from the bottomless cash reserve. Negative A_k indicates the annual investment is moving money from the bottomless cash reserve to S&P. The 0.975, 0.5 and 0.025 quantiles are at the top, middle and bottom respectively. Remaining details about color and points are the same details as in Figure 7.13.

Figure 7.19 shows 0.975, 0.5 and 0.025 quantiles of the total investment, tallied at the end of the year before next year's investment is made. Observe the change in concavity around rate 0.06. 50% of the time, the total investment is quite low compared to the initial investment of 1. As rate increases, so does the total investment.

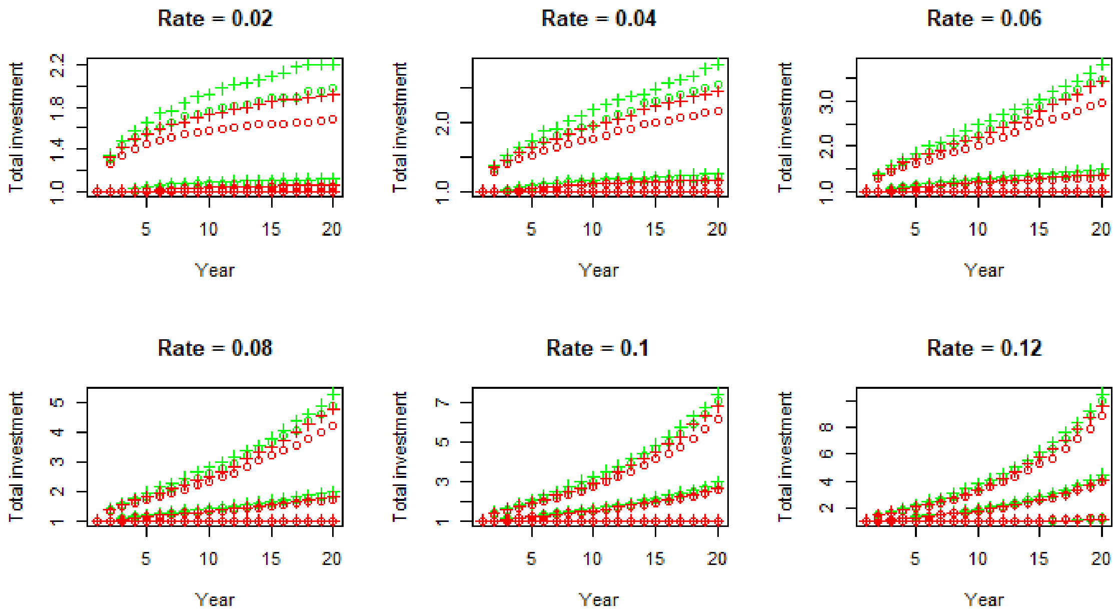

Fig. 7.20: TI_k is plotted for six different rates, where k is the year. The 0.975, 0.5 and 0.025 quantiles are at the top, middle and bottom respectively. Remaining details about color and points are the same details as in Figure 7.13.

7.3.1 *Summary*

Pre-1951 confidence intervals are slightly larger than post-1951 confidence intervals. Using Normal residuals produces larger confidence intervals than kernel density estimates, but the difference is not eggregious. When quantiles are graphed over the proportion, shape is about the same for Normal and kernel density estimates. Overall, the Normal assumption of residuals produces similar results.

Regardless of investment length or rate, real returns are right skewed. There is a rate at which potential returns are maximized. This has to do with the increase in total investment as investment length and rate increase. The investment with the least risk of loss and the investment with the best risk-return tradeoff have a small rate, close to 0. Risk-return tradeoff is better after 1951 than before. Longer investments offer less risk of loss and better risk-return tradeoff.

The annual investment is random, and so is the total investment. Provided a rate less than 0.12 and equal initial investment, the total investment is smaller than in dollar-cost averaging.

8. CONCLUSION

8.1 Dollar-cost averaging for large investment funds

Large investment funds like university endowment funds have certain characterstics that set them apart from most individual investors. In particular, the sheer amount of capital being invested makes investment funds take additional steps to mitigate risk. Special consideration must be given to liquidity and risk of incurring a loss when purchasing assets. For example, investing everything into a single asset can cause problems when the fund managers wish to sell a significant portion of the investment. There may not be enough buyers, and then the investment would be tied up. As will be shown in the following paragraphs, the dollar-cost averaging application of this thesis takes such factors into account, making dollar-cost averaging a viable investing strategy for large investment funds.

First recall that applications use 6-month bonds: US certificates of deposit and US commercial paper to be specific. These are readily available for purchase in large capacities. Moreover, the bonds are insured by the US federal government, making liquidity a non-issue.

In addition to 6-month bonds, applications use the S&P Composite Index. Today, the S&P Composite Index consists of 500 companies, weighted based on market capitalization. As a result of the large amount of capital invested, funds can buy individual stocks at the correct weights, thereby creating a portfolio that follows the S&P Composite Index. Moreover, the fund's capital is dispersed across 500 companies, making it possible to sell a significant portion of the total invested without experiencing major liquidity issues.

Now that the issue of liquidity has been addressed and reduced, the issue of risk is considered. Funds often have many people relying on their success. For example, a unviersity endowment fund cannot take too much risk. If it does, the future of the university can face major setbacks in terms of expansion and renovation. Consequently, funds need reliable returns at manageable risk. The definitions of reliable and manageable vary from fund to fund, so an investing strategy with variable risk-return offerings is needed. Fortunately, the application of dollar-cost averaging presented here provides variable risk-return offerings by tuning the proportion invested in the S&P Composite Index. Moreover, the risk-return offerings are quite favorable, especially in the long-term.

Lastly, a fund needs an investing strategy that is reasonable to implement. Since dollar-cost averaging requires a fixed investment at fixed proportions and time steps, it is predictable and can be planned out. Moreover, it does not require any knack for market timing, so funds can implement it without hiring so-called investment gurus and paying their exorbitant salaries.

Value averaging is not advised for investment funds because of randomness in the periodic investment. It would place a significant burden on a fund to try planning for such randomness. Dollar-cost averaging avoids this randomness with a fixed periodic investment. Note that other options exist for large investment funds besides 6-month US bonds and the S&P Composite Index. For example, 5 year bonds and the Nasdaq 100 Index can be used, with a 5 year time step.

8.2 Final remarks

Two well-known investing strategies were presented: dollar-cost averaging and value averaging. Their recursive definitions led to recursive theoretical results. In the case of dollar-cost averaging, a simpler recursive definition allowed for closed-form representation of some theoretical results. Closed-form results were easily applied. Probability distributions for each strategy were developed recursively, but the integration involved takes too much time and space to apply with accuracy. In addition to the closed-form applications, each strategy was summarized through simulation.

Dollar-cost averaging is attractive because it takes fixed investments at equidistant time steps. It is reasonable for an investor to set aside a fixed amount of income to invest each year. Furthermore, the desired risk-return tradeoff is achieved by adjusting the proportion invested in stocks vs bonds.

The structure of value averaging makes it more attractive in some situations and less in others. Value averaging is less attractive because it takes random investments. Investors must set aside a larger amount, say the 0.95 quantile, even though the actual investment will likely be less. This sort of preparation may not be reasonable for some investors. Value averaging is more attractive because it offers a series of investments that can have a trend other than a straight line. For example, lower rates fit an investor who wants to place a large initial investment, and plans to invest a smaller amount in subsequent years.

Investors looking to invest in S&P and 1YTR can use the results presented earlier to select which strategy and parameters best fits their desired investment amounts, tolerance for risk and desired returns. Investors looking to invest in alternative asset class pairs can modify the code in github [5] to reproduce figures and make investment decisions.

9. APPENDIX

Lemma 5. $R(K) = \frac{1}{4\pi}$ *when K is the bivariate Gaussian kernel.*

Proof. By definition,

$$R(K) = \int_{\mathbb{R}^2} K(x)^2 dx = \int_{\mathbb{R}^2} \left(\frac{1}{2\pi} \exp(-\frac{x_1^2 + x_2^2}{2}) \right)^2 d[x_1 \ x_2]^T). \tag{9.1}$$

The integral can be rewritten as $\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{4\pi^2} \exp(-x_1^2 - x_2^2) dx_1 dx_2$. Continuing to manipulate the integral and noticing the Gaussian integral,

$$\begin{aligned} R(K) &= \frac{1}{4\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} \exp(-x_1^2 - x_2^2) dx_1 dx_2 \\ &= \frac{1}{4\pi^2} \int_{\mathbb{R}} \exp(-x_2^2) \int_{\mathbb{R}} \exp(-x_1^2) dx_1 dx_2 \\ &= \frac{1}{4\pi^2} \int_{\mathbb{R}} \exp(-x_2^2) \sqrt{\pi} dx_2 \\ &= \frac{1}{4\pi^2} \pi \\ &= \frac{1}{4\pi}. \end{aligned} \tag{9.2}$$

□

Lemma 6. $m_2(K) = 1$ *when K is the Gaussian kernel.*

Proof. By definition,

$$m_2(K) I_2 = \int_{\mathbb{R}^2} x x^T K(x) dx = \int_{\mathbb{R}} \int_{\mathbb{R}} \begin{bmatrix} x_1^2 & x_1 x_2 \\ x_1 x_2 & x_2^2 \end{bmatrix} \frac{1}{2\pi} \exp(-\frac{x_1^2 + x_2^2}{2}) dx_1 dx_2. \tag{9.3}$$

Taking the element in position (1,1) of (9.3),

$$\int_{\mathbb{R}} \int_{\mathbb{R}} x_1^2 \frac{1}{2\pi} \exp(-\frac{x_1^2 + x_2^2}{2}) dx_1 dx_2 = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-\frac{x_2^2}{2}) \int_{\mathbb{R}} x_1^2 \exp(-\frac{x_1^2}{2}) dx_1 dx_2 \tag{9.4}$$

Using the substitution $u = -\frac{x_1^2}{2}$ and the Gamma function, the right hand side of (9.4) reduces to

$$\frac{1}{2\pi} \int_{\mathbb{R}} \exp(-\frac{x_2^2}{2}) \sqrt{2\pi} dx_2. \tag{9.5}$$

Noticing the Gaussian integral, (9.5) reduces to

$$\frac{1}{2\pi}\sqrt{2\pi}\sqrt{2\pi} = 1. \tag{9.6}$$

By symmetry, the element in position (2,2) of (9.3) is also 1.

Now to verify that the element in position (1,2) of (9.3) is 0. By symmetry, this will imply that the element in position (2,1) is also 0.

$$\int_{\mathbb{R}}\int_{\mathbb{R}} x_1 x_2 \frac{1}{2\pi}\exp(-\frac{x_1^2+x_2^2}{2})dx_1dx_2 = \frac{1}{2\pi}\int_{\mathbb{R}} x_2\exp(-\frac{x_2^2}{2})\int_{\mathbb{R}} x_1\exp(-\frac{x_1^2}{2})dx_1dx_2 \tag{9.7}$$

Using the substitution $u = \frac{x_1^2}{2}$, the right-hand side of (9.7) reduces to

$$\frac{1}{2\pi}\int_{\mathbb{R}} x_2\exp(-\frac{x_2^2}{2})[\int_0^{\infty} \exp(-u_1)du_1 - \int_0^{\infty} \exp(-u_1)du_1]dx_2 = 0. \tag{9.8}$$

□

BIBLIOGRAPHY

[1] Investors' guide: Dollar cost averaging, 1951.

[2] *THE MONEY MARKETS*, pages 95–130. John Wiley & Sons, Ltd, Chichester, UK, 2018.

[3] Philipp Bagus, Philipp Bagus, Amadeus Gabriel, Amadeus Gabriel, David Howden, and David Howden. On the necessary and sufficient conditions for legitimate banking contracts. *Journal of Business Ethics*, 147(3):669–678, 2018.

[4] Ronald J. Balvers and Douglas W. Mitchell. Autocorrelated returns and optimal intertemporal portfolio choice. *Management Science*, 43(11):1537–1551, 1997.

[5] Hayden Brown. Investing. `https://github.com/HaydenBrown/Investing`, cited October 2020.

[6] Haiwei Chen and Jim Estes. A monte carlo study of the strategies for 401(k) plans: dollar-cost-averaging, value-averaging, and proportional rebalancing. *Financial services review (Greenwich, Conn.)*, 19(2):95, 2010.

[7] James C. Dolley. Dollar cost averaging by industry groups : A stock investment program developed for the endowment fund of the university of texas. *Southwestern social science quarterly*, 48(4):613–623, 1968.

[8] Robert Dubil. The risk and return of investment averaging: an option-theoretic approach. *Financial services review (Greenwich, Conn.)*, 13(4):267, 2004.

[9] Tarn Duong and Martin Hazelton. Plug-in bandwidth matrices for bivariate kernel density estimation. *Journal of Nonparametric Statistics*, 15:17–30, 02 2003.

[10] Michael E. Edleson. *Value Averaging: The Safe and Easy Strategy for Higher Investment Returns*. Wiley, Hoboken, 1. aufl.;1; edition, 2007.

[11] Ming H. Eng and Qing-Guo Wang. Investment strategies for stock markets with mean reversion. pages 2619–2625. IEEE, 2007.

[12] S&P Global. S&p 500. `https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview`, cited October 2020.

[13] Federal Reserve History. Treasury-fed accord. `https://www.federalreservehistory.org/essays/treasury_fed_accord`, cited October 2020.

[14] Martin Širůček and Ivana Škatuĺárová. Application of the value averaging investment method on the us stock market. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 63(6):2151–2160, 2015.

[15] M. Khouja and R. P. Lamb. An optimal schedule for dollar cost averaging under different transaction costs. *International transactions in operational research*, 6(2):245–261, 1999.

[16] J. L. Kirkby, Sovan Mitra, and Duy Nguyen. An analysis of dollar cost averaging and market timing investment strategies. *European journal of operational research*, 286(3):1168–1186, 2020.

[17] Hung-Cheng Lai, Tseng-Chan Tseng, and Sz-Chi Huang. Combining value averaging and bollinger band for an etf trading strategy. *Applied economics*, 48(37):3550–3557, 2016.

[18] Burton G. Malkiel. *A random walk down Wall Street: the time-tested strategy for successful investing*. Norton, New York, 2019.

[19] William T. Morris. Some analysis of purchasing policy. *Management science*, 5(4):443–452, 1959.

[20] Federal Reserve. Data. `https://www.federalreserve.gov/data.htm`, cited October 2020.

[21] William F. Sharpe. Mutual fund performance. *The Journal of business (Chicago, Ill.)*, 39(1):119–138, 1966.

[22] Robert Shiller. Online data robert shiller. `http://www.econ.yale.edu/~shiller/data.htm`, cited October 2020.

[23] 1941-2016 Taylor, Joseph L. *Foundations of analysis*, volume 18. American Mathematical Society, Providence, Rhode Island, 2012.

[24] George W. Wilson. A note on accelerated dollar averaging. *Management science*, 7(2):184–186, 1961.

[25] Jack W. Wilson and Charles P. Jones. An analysis of the s&p 500 index and cowles's extensions: Price indexes and stock returns, 1870–1999. *The Journal of Business*, 75(3):505–533, 2002.

VITA

Hayden Brown is a masters student at the University of Nevada, Reno Department of Mathematics, 2019-2021. His advisor is Andrey Sarantsev. Before this, he completed his undergraduate degrees in Mathematics and Secondary Education at the University of Nevada, Reno, 2015-2019. He is planning to pursue a Ph.D. in Mathematics at the University of Nevada, Reno.