An Analysis Of Dollar Cost Averaging and Market Timing Investment Strategies

J. Lars Kirkby [*] Sovan Mitra [†] Duy Nguyen[‡]

April 28, 2020

Abstract

In this paper we present new theoretical and practical insights into the method of dollar cost averaging (DCA) and averaging-style investment timing strategies, with a formal analysis of the problem. Firstly, we provide a rigorous mathematical formulation for studying DCA and related strategies. This provides closed form formulae for the expected value and variance of the investor's wealth process, which mathematically proves many properties that have been documented in the literature only by empirical studies. Secondly, we prove a counterintuitive, but important, result that the frequency of DCA investment has a non-monotonic and non-trivial impact on risk, risk-return trade-off and other important performance metrics (such as the Sharpe ratio).Thirdly, we provide a method of valuing the DCA risk for models which incorporate jumps. We also provide a method of hedging DCA risk based on applying Asian options. Finally, using the PROJ method of computation, we obtain a robust and computationally efficient method for calculating standard risk measures of generic and deterministic investment strategies, such as DCA. We provide numerical experiments to illustrate our conclusions, and conduct an empirical study on the S&P500 index (from 1954 to 2019) to substantiate our results.

Keywords: dollar cost averaging; market timing; risk management; risk measurement; downside risk; upside risk.

[*]School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30318, USA. Email: jkirkby3@gatech.edu

[†]Department of Mathematics, University of Liverpool, Brownlow Hill, Liverpool, L69 3BX, UK Email: sovan.mitra@liverpool.ac.uk

[‡]Corresponding Author. Department of Mathematics, Marist College, 3399 North Road, Poughkeepsie, NY 12601, USA. Email: nducduy@gmail.com

1 Introduction

Dollar cost averaging (hereafter DCA) is one of the most popular investment methods actively used by ordinary investors. DCA refers to an investment strategy in which a fixed dollar amount is invested in a risky asset (such as a mutual fund or exchange-traded fund (ETF)) at regular time intervals, and over a predefined holding period (see for instance Rubin and Spaht (2018), Panyagometh and Zhu (2016) and Fong (2017)). The commonly publicized advantage of this form of incremental investing is that the investor reduces exposure to temporary price fluctuations, which could lead to lower returns. For example, during a market downturn one is able to buy more shares of the asset at a lower price, leading to greater returns at the end of the investment horizon. By reducing exposure to the impact of price fluctuations, DCA enables investors to avoid "market timing". Hence DCA has been found in some studies to achieve better investment returns for ordinary investors, without requiring sophisticated market timing skills[1]. Additionally, DCA is a standard method of saving for many important investments for future wealth, such as retirement accounts and pensions. Hence it is important to understand DCA as an investment method. In general, asset/portfolio allocation and investment timing is an important problem, with many practical applications, Wachter (2010); Battauz et al. (2015, 2017); Battauz and Sbuelz (2018); Ma et al. (2019); Pun and Wong (2019); Ling et al. (2019).

The advantage of DCA not requiring market timing skills is particularly appealing, especially for ordinary investors, as it has been well documented that market timing is extremely difficult to achieve, even for professional or experienced investors. For example, Cuthbertson et al. (2010) show that only approximately 1% of UK mutual funds demonstrate significant market timing skills. Economically, this implies that learning (costs) in terms of the optimal investment timing are extremely high, and so DCA provides a method of bypassing the learning of such skills. Additionally, DCA is preferable to lump sum (LS) investing from an economic perspective because DCA places fewer liquidity and other budgetary constraints upon firms and individuals (see for instance Telyukova (2013), Campbell and Hercowitz (2019) and Boguth and Simutin (2018)).

Another potential advantage of DCA is that it is deemed to reduce the volatility of investments (Braselton et al., 1999), by smoothing out the effects of volatile movements in asset prices with regular, periodic asset purchases. This has also been referred to as *time diversification*, in that it reduces the risk from purchasing at the "incorrect" time. Under suitable assumptions, we prove that DCA does in fact offer a reduction in variance compared with LS investing, but

[1] In Tomlinson (2012). DCA is deemed an "the unbeatable formula" because "No one has yet discovered any other formula for investing which can be used with so much confidence of ultimate success, regardless of what may happen to security prices, as Dollar Cost Averaging." While this is a positive appraisal of DCA, we will investigate the potential merits of DCA and related strategies from a mathematical and formal perspective in this paper.

at the cost of lower expected returns.

Nonetheless, academic views on DCA have been generally negative. For example, Constantinides (1979) discredited the efficiency of DCA as it is considered to be an inferior strategy. His conclusion was mainly drawn on DCA not having flexibility to incorporate additional information that is accumulated as time passes; DCA is a *deterministic policy* that is not adjusted as new market information is revealed. In Knight and Mandell (1992), using graphical analysis, historical stock market returns, and Monte Carlo simulation, the authors argued that DCA is suboptimal. In an empirical study, Williams and Bacon (1993) compares the annualized returns from various dollar-cost averaging strategies with those produced by lump-sum (LS) investing in the S&P 500 from 1926 to 1991. For all time periods and averaging strategies investigated, LS investing produced superior returns to DCA, and in all but one instance, the differences were significant at the 0.005 level. In an empirical study with S&P 500 data, as well as simulated experiments, we will show that a more optimal *deterministic* strategy is generally somewhere "between" DCA and LS investing, but it depends critically on the frequency of periodic investments, and on the investment horizon.

Despite the aforementioned unfavorable views on DCA, there is also substantial (empirical) research evidence in support of DCA. For example, Trainor (2005); Dubil (2005) showed that DCA reduces shortfall risk, which is a major factor in retirement saving and retail investing. Grable and Chatterjee (2015) documented that a DCA strategy provides a way to outperform during a bear market (rather than a bull market) for clients who have less tolerance for financial risk. Luskin (2017) reported that DCA outperforms LS investment during the period when cyclically adjusted price-to-earnings (CAPE) ratio is high. Balvers and Mitchell (2000); Brennan et al. (2005) argued that DCA may be advantageous if asset returns are negatively autocorrelated (e.g. correcting). Moreover, using mutual fund data, Israelsen (1999) argued that lump-sum investing does not always yield superior returns over dollar-cost averaging, especially if the volatility of mutual funds is low. For additional literature on DCA and LS investing, the reader is referred to Pye (1971); Dodson (1989); Thorley (1994); Leggio and Lien (2003); Milevsky and Posner (2003); Smith and Artigue (2018) and references therein.

Whilst DCA has several important advantages for ordinary investors, there has been little research conducted with respect to continuous time financial models, which are fundamental to financial risk and derivatives modeling. For example, under the Geometric Brownian motion framework, Milevsky and Posner (2003) prove that the expected return from the DCA strategy, conditional on knowing the final value of the security, will uniformly exceed the return from the underlying security for all sufficiently large volatilities. Vanduffel et al. (2012) revisited the suboptimality of DCA when (log)returns are governed by Lévy processes. They then construct a strategy that dominates the DCA strategy explicitly. In this work, we consider the case of general exponential Lévy models, which includes the classic Black-Scholes-Merton framework, but also allows for jumps in the underlying.

1.1 Summary of Contributions

In this paper we present new theoretical and practical insights into the nature of DCA and deterministic investment timing strategies in general. We provide a formal and rigorous mathematical formulation for examining DCA and related strategies, which confirms many of the DCA properties that have been documented in the literature only by empirical studies. Given the importance of risk measures such as VaR and Expected Shortfall in determining optimal investment strategies in numerous applications Staino and Russo (2019); Ling et al. (2019), we provide an efficient method for calculating such measures under DCA and deterministic strategies in general. Since the seminal work of Markowitz (1952), mean-variance optimality is another commonly used method for portfolio selection Lwin et al. (2017); Bi et al. (2018); Penev et al. (2019), and we also provide the mean-variance optimal deterministic strategy.

The contributions of this paper are as follows:

- Firstly, we provide a rigorous mathematical formulation for studying deterministic timing strategies, such as DCA and geometric DCA. Fundamental results are derived, which confirm many empirical properties that have been reported in the literature. Several counter-intuitive results are also derived and verified empirically. We then derive in closed-form the mean-variance optimal timing strategy. The mathematical results are supported by numerous examples, which illustrate the relationship between the various investment strategies, and provide insights into the risk-return tradeoffs afforded by each strategy.

- Secondly, we provide a method of evaluating DCA risk (and deterministic investment timing strategies in general) for continuous time models which incorporate jumps. We provide a unified framework for modeling and analyzing the risk of DCA, and for determining optimal deterministic strategies based on various risk measures. We prove a fundamental distributional equality between timing strategies and arithmetic averaging, which enables the use of efficient computational methods for risk measurement and hedging. From this, we develop an efficient and robust numerical method based on the Fourier PROJ method for evaluating the risk of deterministic strategies. We also provide numerical results which can be used as a reference for industry and research work.

- Finally, we conduct an empirical study using S&P500 index (SPX) to assess the merits of LS, DCA, and a generalization of the two, which we call Geometric DCA, under various investment horizons. The study not only confirms a number of theoretical results presented in the paper but also provides new insights on how one can use a Geometric DCA investment strategy to obtain a better tradeoff between risk and return.

The rest of the paper is organised as follows. In Section 2, we introduce the exponential Lévy processes which are considered as models for the risky investment. In Section 3, we formally define the deterministic timing problem in full generality, and then discuss the important special cases of DCA and Lump Sum (LS) investing, as well as the relationships between these common investment strategies. In particular we examine the frequency of investment, risk-return tradeoffs, and related performance metrics (such as the Sharpe ratio). In Section 4, we introduce the PROJ computational method for calculating risk measures on the wealth outcome of timing strategies (such as Value at Risk). In Section 5, we conduct an empirical study to assess the merits of DCA, Geometric DCA, and LS investment strategies applied to the S&P500 index (SPX) over the period 1954-07-01 to 2019-04-11. In this section we compare several metrics, including Sharpe's ratio, CRRA utility, as well as the mean and variance of terminal wealth for several competing timing strategies. We finally end with a conclusion.

2 Exponential Lévy process

We formulate the DCA investment problem for investing in a risky asset (mutual fund, exchange-traded fund (ETF), index, stock, etc.) using a class of Lévy models which are common in financial asset pricing and risk modeling. Let $\{L(t)\}_{t\geq 0}$ be a Lévy process, a continuous-time stochastic process, with stationary and independent increments, which generalizes the classic diffusion process by allowing for jumps. We assume that our random processes are defined on the standard filtered probability space $(\Omega; \mathcal{F}; P)$, and $\mathcal{F}_t, t \geq 0$, represents the information available at time t. Many stochastic processes are included in this class, the most well known being the scaled Brownian motion with drift Bachelier (1900) where $L_{t+\Delta t} - L_t \stackrel{d}{=} \mathcal{N}((\mu - \frac{\sigma^2}{2})\Delta t, \sigma^2 \Delta t)$. Here "$\stackrel{d}{=}$" is used to denote equality in distribution, and $\mathcal{N}(\mu, \sigma^2)$ denotes a normally distributed random variable with mean μ and standard deviation σ. To ensure positivity in asset prices, the Geometric Brownian Motion (GBM) model of Black and Scholes (1973) posited the dynamics

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \quad 0 \leq t \leq T. \tag{1}$$

Jumps can be added to the GBM model which leads to Merton's jump diffusion (MJD) model, see Merton (1976), given by

$$dS_t = \mu S_{t^-} dt + \sigma S_{t^-} dB_t + S_{t^-} d\left(\sum_{i=1}^{N_t} (Y_i - 1)\right), \tag{2}$$

where N_t is a compound Poisson process with rate $\lambda > 0$. In the MJD model, the jump size Y_i satisfies $\log(Y_i) \sim \mathcal{N}(\mu_J, \sigma_J^2)$. Another common model for $\log(Y_i)$ is when $\log(Y_i)$ has an asymmetric double exponential distribution, with density

$$f(z) = p\eta_1 e^{-\eta_1 z}\mathbb{1}_{\{z\geq 0\}} + (1-p)\eta_2 e^{\eta_2 z}\mathbb{1}_{\{z<0\}}, \quad 0 < p < 1, \eta_1 > 1, \eta_2 > 0,$$

where $\mathbb{1}_{\{.\}}$ denotes the indicator function, then (2) is reduced to Kou's jump diffusion model (see Kou (2002)). More generally, we consider the class of *exponential Lévy models*,

$$S_t = S_0 e^{\mu t + L(t)}, \quad t \geq 0, \tag{3}$$

where μ is the expected annualized rate of return on the asset. From the Lévy-Khintchine theorem, the characteristic function (ChF) of $L(t)$ satisfies

$$\phi_{L(t)}(\xi) := \mathbb{E}[e^{iL(t)\xi}] = e^{t\psi_L(\xi)}, \quad t \geq 0. \tag{4}$$

In the above equation, $\psi_L(\xi)$ is called the *Lévy symbol* of $L(t)$. The model considered in (3) covers a wide range of common exponential Lévy processes. In particular, it covers the following family of processes:

- *Carr-Geman-Madan-Yor (CGMY)*. The CGMY family of exponential Lévy models introduced in Carr et al. (2002) is characterized by four parameters: $C > 0$ accounts for the overall activity level, $G > 0$ and $M > 0$ control the distribution's skewness, and $0 < Y < 2$ dictates the fine structure of the process. This model is a popular parameterization of the more general KoBoL class of Boyarchenko and Levendorskii (2000, 2002).

- *Normal Inverse Gaussian (NIG)*. The $NIG(\alpha, \beta, \delta)$ return process Y_t of Barndorff-Nielsen (1997) is constructed by time changing a standard Brownian motion, via the inverse Gaussian subordinator I_t which has parameters $a = 1$ and $b = \delta\sqrt{\alpha^2 - \beta^2}$, $Y_t = \beta\delta^2 I_t + \delta W_t$, where $\alpha > 0, \beta \in (-\alpha, \alpha), \delta > 0$.

- *Variance Gamma (VG)*. The VG process introduced in Madan et al. (1998) has finite moments which distinguishes it from many Lévy processes and is also determined by a random time change. The process can be written as a Brownian motion with drift subordinated by a random time change which follows a Gamma process.

For a list of characteristic functions of several common exponential Lévy processes, please see Table 3 in Appendix A. Throughout, we will assume that all investments are made in one of two accounts. The first is the random asset (fund) S_t, which is modeled as an exponential Lévy process in (3). In particular, it is assumed that $L(t)$ is normalized such that $\mathbb{E}[e^{L(t)}] = 1$, which implies $\mathbb{E}[S_{t+\Delta t}|S_t] = S_t e^{\mu\Delta t}$ for $\Delta t, t \geq 0$.[2] In particular, we will model the process using the *convexity-corrected symbol*[3]

$$\psi_R(\xi) = i(\mu - \psi_L(-i))\xi + \psi_L(\xi),$$

which ensures that $\mathbb{E}[e^{L(t)}] = 1$ for any $\psi_L(\xi)$. For example, in the GBM case (1), we have $S_t = S_0 e^{\mu t + L(t)} = S_0 e^{(\mu-\sigma^2/2)t+\sigma B(t)}$. In particular, the Lévy process $L(t) = -(\sigma^2/2)t + \sigma B(t)$ includes the convexity correction, which satisfies the requirement $1 = \mathbb{E}[e^{L(t)}] = e^{-(\sigma^2/2)t}\mathbb{E}[e^{\sigma B(t)}]$.

The second asset is a risk free account, with a guaranteed annualized rate of return of $r \in \mathbb{R}$, where we assume that $\mu \geq r$.

In this Lévy market, each investor has her/his own utility function $u(\cdot)$ which is used to assign a utility $u(w)$ to each possible wealth level w. When the investor must choose between two random wealth outcomes X and Y, she/he compares the expected utilities, $\mathbb{E}[u(X)]$ and $\mathbb{E}[u(Y)]$, provided that these expectations exist. An investor is said to be *profit seeking* if her/his utility function $u(\cdot)$ in non-decreasing, see Vanduffel et al. (2012) for more discussions related to the present context.

Throughout this paper, it is assumed that every reasonable investor prefers a certain gain over a random gain with the same expectation. This in turn implies that her/his utility function is concave, that is

$$\mathbb{E}[u(X)] \leq u(\mathbb{E}[X]). \tag{5}$$

We assume that investors are *risk averse*, i.e. their utility function $u(\cdot)$ is concave and non-decreasing. Lastly, to ensure economic consistency of Lévy processes, it is assumed throughout this paper that the expected return on the risky investment is never less than that of the risk free investment, that is $\mathbb{E}[S_T|S_t] \geq e^{r(T-t)}$. In the Lévy market, this is equivalent to the assumption that $\mu \geq r$.

3 Discrete Investment Formulation

We begin by formulating the general problem of an investor who seeks to optimally invest a fixed wealth $W = W_0$ into an asset over a time horizon $[0, T]$, such that all wealth is allocated by T. The initial endowment currently resides in an account which earns the risk-free rate of interest[4], $r \in \mathbb{R}$. The trajectory of the asset (index) price is denoted $(S_t)_{t\geq 0}$, and we assume that the asset earns an expected rate of return of $\mathbb{E}[S_{t+\Delta t}/S_t|S_t] = \exp(\mu\Delta t)$, where $\mu \geq r$.

[2]Note that under the physical measure, μ needn't equal the risk free rate of interest, so S_t is not assumed to be a martingale after discounting.

[3]Note that we refer to convexity here in the mathematical function sense rather than with respect to the pure financial terminology (such as bond convexity correction).

[4]Note that interest rates can be negative in this analysis, which is the current market state within the Eurozone, for example.

Let $\alpha_m \geq 0$, $\alpha_m \in \mathcal{F}_0$, denote the dollar value invested at any one of the discrete times $0 = t_0 < t_1 < \cdots < t_M = T$ representing the beginning of the investment period $[t_m, t_{m+1}]$, at a price of S_{t_m}, where $m = 0, 1, 2, ..., M-1$. At the beginning of each stage t_m, the investor purchases $n_m := \alpha_m / S_{t_m}$ shares of the asset, leaving an uninvested cash holding of $W - \sum_{j=0}^{m} \alpha_j$, which earns a risk-free income over the investment period $[t_m, t_{m+1}]$ of

$$\left(e^{r(t_{m+1}-t_m)} - 1\right)\left(W - \sum_{j=0}^{m} \alpha_j\right), \quad m = 0, \ldots, M-1,$$

where continuous compounding is assumed. Moreover, interest also accrues on the cash flows generated by previous risk-free investment holdings. In the final stage, any remaining wealth α_M (from the initial endowment of W) is invested. For any *deterministic timing strategy*[5] $\boldsymbol{\alpha} = \{\alpha_m\}_{m=0}^{M} \in \mathcal{F}_0$, we can thus express the terminal wealth, $W_T = W_T(\boldsymbol{\alpha})$, as the sum of two accounts:

$$W_T(\boldsymbol{\alpha}) = W_T^S(\boldsymbol{\alpha}) + W_T^R(\boldsymbol{\alpha}),$$

where the first term on the right hand side is the wealth invested in the risky asset, and the second term is the wealth resulting from all risk-free proceeds up to T. In Lemma 1, we give an explicit representation of $W_T(\boldsymbol{\alpha})$.

Lemma 1. *Let $\boldsymbol{\alpha} = \{\alpha_m\}_{m=0}^{M} \in \mathcal{F}_0$, $M \geq 1$, be an investment strategy with initial endowment W such that $\alpha_m \geq 0$ and $\sum_{m=0}^{M} \alpha_m = W$. The terminal wealth $W_T(\boldsymbol{\alpha}) \in \mathcal{F}_T$ satisfies*

$$\begin{aligned} W_T(\boldsymbol{\alpha}) &= S_T \cdot \sum_{m=0}^{M} \frac{\alpha_m}{S_{t_m}} + e^{rT} W - \sum_{m=0}^{M} \alpha_m e^{r(T-t_m)} \\ &= S_T \cdot \sum_{m=0}^{M-1} \frac{\alpha_m}{S_{t_m}} + e^{rT} W - \sum_{m=0}^{M-1} \alpha_m e^{r(T-t_m)}. \end{aligned} \tag{6}$$

Proof. First, since the total shares purchased at the end is $\sum_{m=0}^{M} n_m = \sum_{m=0}^{M} \frac{\alpha_m}{S_{t_m}}$, it is clear that $W_T^S(\boldsymbol{\alpha}) = S_T \cdot \sum_{m=0}^{M} \frac{\alpha_m}{S_{t_m}} \in \mathcal{F}_T$. We will prove that $W_T^R(\boldsymbol{\alpha}) = e^{rT} W - \sum_{m=0}^{M} \alpha_m e^{r(T-t_m)} \in \mathcal{F}_0$. To this end, let us denote by $W_{t_m}^R$ the risk-free account value at the end of period t_m, that is immediately following the investment of size α_m in the asset, then we have the update rule

$$W_{t_m}^R = e^{r(t_m - t_{m-1})} W_{t_{m-1}}^R - \alpha_m, \quad m = 1, \ldots, M,$$

where $W_{t_0}^R = W - \alpha_0$, which is the sum of interest gains and what remains of the initial wealth

[5]That is, any strategy which is determined at time t_0, and adhered to over $[0, T]$.

W to be invested. Hence

$$
\begin{aligned}
W_{t_m}^R &= e^{r(t_m - t_{m-1})} W_{t_{m-1}}^R - \alpha_m \\
&= e^{r(t_m - t_{m-1})} \underbrace{\left(e^{r(t_{m-1} - t_{m-2})} W_{t_{m-2}}^R - \alpha_{m-1}\right)}_{W_{t_{m-1}}^R} - \alpha_m \\
&= \dots \\
&= \prod_{p=1}^{m} e^{r(t_p - t_{p-1})} W - \sum_{k=0}^{m} \prod_{p=k+1}^{m} e^{r(t_p - t_{p-1})} \alpha_k \\
&= e^{r(t_m - t_0)} W - \sum_{k=0}^{m} \alpha_k e^{r(t_m - t_k)} = e^{rT} W - \sum_{k=0}^{M} \alpha_k e^{r(T - t_k)} \quad \text{when} \quad m = M,
\end{aligned}
$$

where $t_0 = 0$ and $t_m = T$. Hence

$$
\begin{aligned}
W_T(\boldsymbol{\alpha}) &= W_T^S(\boldsymbol{\alpha}) + W_T^R(\boldsymbol{\alpha}) \\
&= S_T \cdot \sum_{m=0}^{M} \frac{\alpha_m}{S_{t_m}} + e^{rT} W - \sum_{m=0}^{M} \alpha_m e^{r(T - t_m)},
\end{aligned}
$$

and the result follows immediately. $\square$

In general, the decision is how to choose $\boldsymbol{\alpha} = \{\alpha_m\}_{m=0}^{M}$ to optimize the utility of this wealth (or to minimize some measure of risk). That is, invest the initial endowment optimally by choosing $\boldsymbol{\alpha}$ to solve

$$
\max_{\boldsymbol{\alpha}} u\left(W_T(\boldsymbol{\alpha})\right) \quad \text{s.t.} \quad \alpha_m \geq 0, \quad \sum_{m=0}^{M} \alpha_m = W, \tag{7}
$$

where $u : \mathbb{R} \to \mathbb{R}$ captures the (expected) utility of the (random) terminal wealth. We note that u can represent a standard utility function, or an objective measure such as Sharpe's ratio or other mean-variance criteria. For example, a common strategy is to maximize $u(W_T(\boldsymbol{\alpha})) = \theta \mathbb{E}\left[W_T(\boldsymbol{\alpha})\right] - Var\left(W_T(\boldsymbol{\alpha})\right)$, where θ controls the level of risk-aversion. Due to the strong path dependency of the deterministic timing problem, the optimal solution is extremely difficult to determine for non-trivial u, and heuristics (such as DCA) are attractive approaches.[6]

Given the importance of mean and variance for quantifying and comparing risky outcomes, we next derive these expression for a general deterministic timing strategy.

Lemma 2. *Suppose that the expected asset growth satisfies* $\mathbb{E}[S_{t+\Delta t}/S_t | S_t] = \exp(\mu \Delta t)$, *for* $t, \Delta t \geq 0$. *Then*

$$
\mathbb{E}\left[W_T(\boldsymbol{\alpha})\right] = e^{rT} W + \sum_{m=0}^{M-1} \alpha_m \left(e^{\mu(T - t_m)} - e^{r(T - t_m)}\right). \tag{8}
$$

[6]For the remaining discussion, the reader can keep in mind the identity utility, $u(x) = x$. A more interesting utility function is considered in Section 5.

If $\mu \geq r$, the expected terminal wealth of any strategy $\boldsymbol{\alpha} = \{\alpha_m\}_{m=0}^M \in \mathcal{F}_0$ can be bounded by

$$e^{rT} W \leq \mathbb{E}[W_T(\boldsymbol{\alpha})] \leq e^{\mu T} W. \tag{9}$$

If $\mu > r$, so that the risky investment asset offers greater expected returns than the risk free rate, the upper bound is achieved by the lump-sum strategy $\alpha_0 = W$, and the lower bound by $\alpha_M = W$. In this case, for any strategy $\boldsymbol{\alpha} = \{\alpha_m\}_{m=0}^M$ with $\alpha_M < W$, we have $\mathbb{E}[W_T(\boldsymbol{\alpha})] > W \cdot \exp(rT)$.

Proof. See appendix B. □

From Lemma 2 if $\mu > r$, then an investor who seeks to simply optimize the expected terminal wealth, $u(W_T(\boldsymbol{\alpha})) = \mathbb{E}[W_T(\boldsymbol{\alpha})]$, or equivalently, $e^{-rT}\mathbb{E}[W_T(\boldsymbol{\alpha})]$, will always prefer a lump sum investment of $\alpha_0 = W$, which earns an expected return of $\mathbb{E}[W_T(\boldsymbol{\alpha})/W] - 1 = \exp(\mu T) - 1 > \exp(rT) - 1$. This proves that in this case DCA is suboptimal (excluding the risk aversion of the investor).[7]

In addition to the final wealth, investors are also concerned about the uncertainty of their wealth, which can be measured, for example, by the variance. At one extreme, the strategy with $\alpha_M = W$ results in a strategy with zero variance of W_T. In general, we have the following variance of wealth for any deterministic strategy.

Lemma 3. *Suppose that S_t is an exponential Lévy process defined in* (3)*, with symbol $\psi_R(\xi) = i(\mu - \psi_L(-i))\xi + \psi_L(\xi)$. Then, for $M \geq 1$, $\boldsymbol{\alpha} = \{\alpha_m\}_{m=0}^M \in \mathcal{F}_0$,*

$$\begin{aligned} Var\left(W_T(\boldsymbol{\alpha})\right) = &\sum_{m=0}^{M-1} \alpha_m^2 \Big\{ \exp((M-m)\Delta t \cdot \psi_R(-2i)) - \exp(2(M-m)\Delta t \cdot \mu) \Big\} \\ &+ 2 \sum_{m=0}^{M-2} \sum_{k=1}^{M-m-1} \alpha_m \alpha_{m+k} \Big\{ \exp(k\Delta t \cdot \mu) \exp((M-m-k)\Delta t \cdot \psi_R(-2i)) \\ &\qquad - \exp((2(M-m)-k)\Delta t \cdot \mu) \Big\}. \end{aligned} \tag{10}$$

Proof. See appendix B. □

We note that for exponential Lévy processes, the symbol $\psi_R(\xi) = i(\mu - \psi_L(-i))\xi + \psi_L(\xi)$ is typically known explicitly. In Table 3 in Appendix A, we provide the closed functional form $\psi_R(\xi)$ and its value $\psi_R(-2i)$ for several Lévy processes, from which (10) is calculated in closed-form.

[7] When $r = \mu$, all strategies are equivalent, and expected wealth alone is not sufficient to prefer a strategy.

3.1 DCA Fundamental Results

The first deterministic investment strategy we investigate is DCA, where in each period we invest an identical amount $\alpha_m = W/(M+1)$ for $m = 0, \ldots, M$, and (6) becomes

$$\begin{aligned} W_T^{DCA} &= We^{rT} + \frac{W}{M+1} \sum_{m=0}^{M} \left(\frac{S_T}{S_{t_m}} - e^{r(T-t_m)} \right) \\ &= \frac{W}{M+1} \sum_{m=0}^{M} \left(\frac{S_T}{S_{t_m}} + e^{rT} \left(1 - e^{-rt_m}\right) \right). \end{aligned}$$

We next derive several fundamental results concerning DCA investing. For example, we demonstrate that the wealth variance is non-trivial, and depends significantly on M. This is a counter-intuitive, but also an important result, because M is typically neglected in analyses of deterministic strategies.

Recall that the number of shares purchased at time t_m is determined by $n_m = \alpha_m / S_{t_m} = \frac{W}{(M+1)S_{t_m}}$. As a result, an investor who uses a DCA strategy must (periodically) adjust the number of shares purchased as asset prices fluctuate. If asset prices S_{t_m} are high, n_m is small, so fewer shares are bought, and conversely the same is also true. Since the average cost weights the purchase prices by the number of shares acquired at each price, the average cost is always less than the average price. This implies that DCA provides a better (overall) purchasing price compared to the average asset price. Specifically, we have the following theorem:

Theorem 1. *For a DCA strategy, the average purchase cost is always less than or equal to the average asset price. That is, for $M \geq 1$, and $\alpha_m = W/(M+1)$ for $m = 0, \ldots, M$, we have*

$$\frac{W}{n_0 + n_1 + \ldots + n_M} \leq \frac{S_{t_0} + S_{t_1} + \ldots + S_{t_M}}{M+1}.$$

Proof. See appendix B. □

From Theorem 1, it appears that a DCA strategy seems to have a positive expected profit even if changes in asset prices are independent with a mean of zero. However, the fact that the expected value of average cost is less than the expected value of the asset prices does not always imply a positive expected return (see Smith and Artigue (2018)). The following corollary gives us a surprising result reflecting the expected performance of DCA as a function of M, in that the expected value decreases with M.

Corollary 2. *Suppose that $\mathbb{E}[S_{t+\Delta t}/S_t | S_t] = \exp(\mu \Delta t)$, for $t, \Delta t \geq 0$, for example when S_t is an exponential Lévy process, and $t_m = Tm/M$. Then the expected value $\mathbb{E}\left[W_T^{DCA}\right]$ is given by*

$$\mathbb{E}\left[W_T^{DCA}\right] = e^{rT} W + \frac{W}{M+1} \left(\frac{e^{\mu T(M+1)/M} - 1}{e^{\mu T/M} - 1} - \frac{e^{rT(M+1)/M} - 1}{e^{rT/M} - 1} \right), \quad M \geq 1,$$

and $\mathbb{E}\left[W_T^{DCA}\right]$ *is monotonically decreasing in* M*. Moreover we have for* $r \neq 0, \mu \neq 0$,

$$\lim_{M\to\infty} \mathbb{E}\left[W_T^{DCA}\right] = e^{rT}W + W\left(\frac{e^{\mu T}-1}{\mu T} - \frac{e^{rT}-1}{rT}\right). \tag{11}$$

Proof. See appendix B. □

In particular, expected terminal wealth *always* declines as we increase the number of investments between t_0 and T, eventually approaching the limit in (11). As a corollary of Lemma 3, we can obtain a closed-form for the variance of the DCA investment strategy.

Corollary 3. *The variance of DCA under exponential Lévy dynamics for* $M \geq 1$ *is given in closed form by*

$$Var\left[W_T^{DCA}\right] = \frac{W^2}{(M+1)^2}\left(\Pi_1 + 2\Pi_2 - 2\Pi_3\right), \tag{12}$$

where Π_1, Π_2, Π_3 *are defined as*

$$\Pi_1 := \frac{\exp((M+1)\Delta t\cdot\psi_R(-2i))-1}{\exp(\Delta t\cdot\psi_R(-2i))-1} - \frac{\exp(2(M+1)\Delta t\cdot\mu)-1}{\exp(2\Delta t\cdot\mu)-1}$$

$$\Pi_2 := \frac{\exp(\Delta t\cdot(\mu-\psi_R(-2i)))}{\exp(\Delta t(\mu-\psi_R(-2i)))-1}\cdot\left(\exp(\Delta t(\mu+\psi_R(-2i)))\frac{\exp((M-1)\Delta t\cdot\mu)-1}{\exp(\Delta t\cdot\mu)-1}\right.$$
$$\left. - \exp(2\cdot\Delta t\cdot\psi_R(-2i))\frac{\exp((M-1)\Delta t\cdot\psi_R(-2i))-1}{\exp(\Delta t\cdot\psi_R(-2i))-1}\right).$$

$$\Pi_3 := \frac{\exp(2\Delta t\cdot\mu)}{\exp(-\Delta t\cdot\mu)-1}\left(\frac{\exp((M-1)\Delta t\mu)-1}{\exp(\Delta t\mu)-1} - \exp(\Delta t\cdot\mu)\frac{\exp(2(M-1)\Delta t\cdot\mu))-1}{\exp(2\Delta t\cdot\mu)-1}\right).$$

Proof. See appendix B. □

We note that in contrast to its expectation, the variance of DCA is not always monotonic in M. To demonstrate the return-uncertainty trade-off, we consider the terminal wealth under a Geometric Brownian motion (GBM) model for S_t. The underlying has a volatility of $\sigma = 0.2$, which is typical of equities, with an expected return (drift) of $\mu = 0.08$. The risk-free asset delivers a continuously compounded return of $r = 0.02$, and the investor starts with $W = 10$ units of wealth which is invested over $M+1$ periods. We also consider a Merton jump diffusion (MJD) model (recall (2)), with a high volatility $\sigma = 0.6$, and a high jump rate of $\lambda = 6$.

In Figure 1, we illustrate the theoretical mean/variance of DCA under the two models, over an investment horizon of $T = 1$ year, as a function of M. As expected, from the plots in Figure 1 the expectation (mean) decreases monotonically to the limit in (11) as M increases while the variance is not always monotonic. However, the variance tends to increase for large M (in fact the variance converges to its limit derived in Corollary 4). The figure illustrates that the "optimal" investment strategy is not trivial, and it will depend crucially on the choice of M, which governs the number of investments made over $[0, T]$. Surprisingly, for the (non-pathological) example in the left figure, any natural trade-off between return and risk will favor

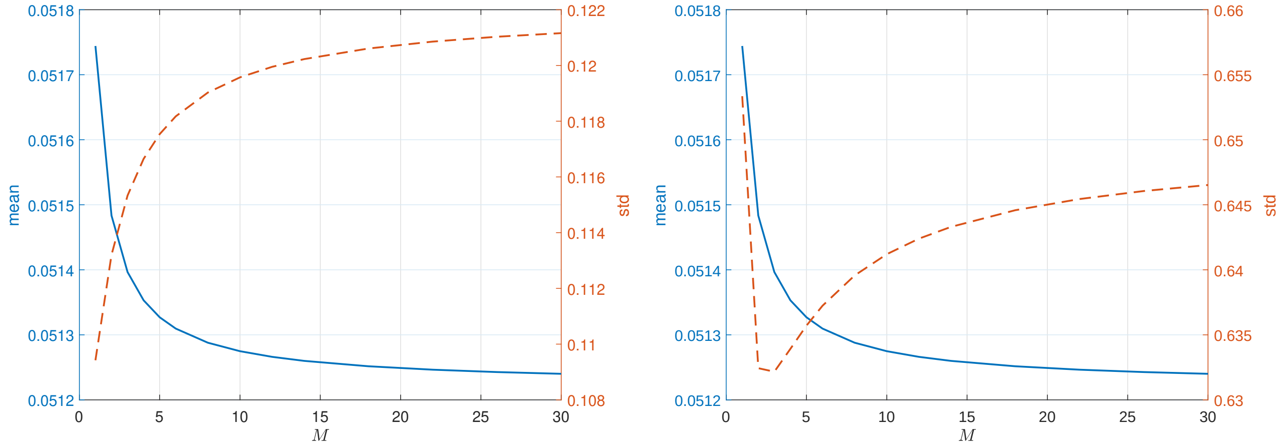

Figure 1: DCA mean (solid line) and std. dev. (dashed line) of return on initial wealth, $(W_T - W)/W$, for DCA as a function of M, with $r = 0.02$, $\mu = 0.08$, $T = 1$, $W = 10$. Left: GBM with $\sigma = 0.2$. Right: MJD, $\sigma = 0.6, \lambda = 6, \mu_J = -0.2, \sigma_J = 0.3$.

$M = 1$ for DCA, which represents an equal split of wealth between $t = 0$ and $t = T$. In the previous examples, we saw that $Var(W_T^{DCA})$ is not always a monotone function of M. It is however bounded, and it converges to a limit.

Corollary 4. *The variance of wealth under DCA investing converges in the limit to*

$$\lim_{M\to\infty} Var\left[W_T^{DCA}\right] = \frac{2W^2}{T(\mu - \psi_R(-2i))}\left(\frac{\exp(T\mu) - 1}{T\mu} - \frac{\exp(T\psi_R(-2i)) - 1}{T\psi_R(-2i)}\right) + \frac{W^2}{(T\mu)^2}\left(2\exp(T\mu) - \exp(2T\mu) - 1\right). \tag{13}$$

Proof. See appendix B. ☐

These results offer precious insight into DCA investing, proving that different choices of M lead to different levels of mean, variance, and equally as important risk-return trade-offs, and hence to different optimal investment strategies.

3.2 Comparison to Lump Sum Investing

We next compare DCA to another common deterministic timing strategy, lump sum (LS) investing, to illustrate the trade-off between return and uncertainty of terminal wealth. In the case of LS investing, we invest the full amount W at some period $\widetilde{m} \in \{0, 1, \ldots, M\}$, so

$$W_T^{LS_{\widetilde{m}}} = W\left(\frac{S_T}{S_{t_{\widetilde{m}}}} + e^{rT}\left(1 - e^{-rt_{\widetilde{m}}}\right)\right), \tag{14}$$

where we denote this strategy by $LS_{\widetilde{m}}$. Most commonly, $\widetilde{m} = 0$.

Intuitively, the uncertainty of a timing strategy is a measure of the *duration* of time that the wealth is exposed to the risky asset. This is illustrated clearly in Figure 2, which compares the same DCA strategy to each of the possible LS strategies, which differ based on which time period, $t_{\widetilde{m}}$ for $\widetilde{m} \in \{0, \ldots, M\}$, is chosen to invest the full wealth. From (14), the investment wealth and risk can be calculated easily from the single source of randomness, $S_T/S_{t_{\widetilde{m}}}$, which yields

$$\mathbb{E}\left[W_T^{LS_{\widetilde{m}}}\right] = e^{rT}W + W\left(e^{\mu(T-t_{\widetilde{m}})} - e^{r(T-t_{\widetilde{m}})}\right), \tag{15}$$

$$Var(W_T^{LS_{\widetilde{m}}}) = W^2\big(\exp((T-t_{\widetilde{m}})\psi_R(-2i)) - \exp((T-t_{\widetilde{m}})2\mu)\big).$$

In particular, both the expected return and variance are monotonically decreasing in $\widetilde{m}$, which is seen as well in Figure 2.

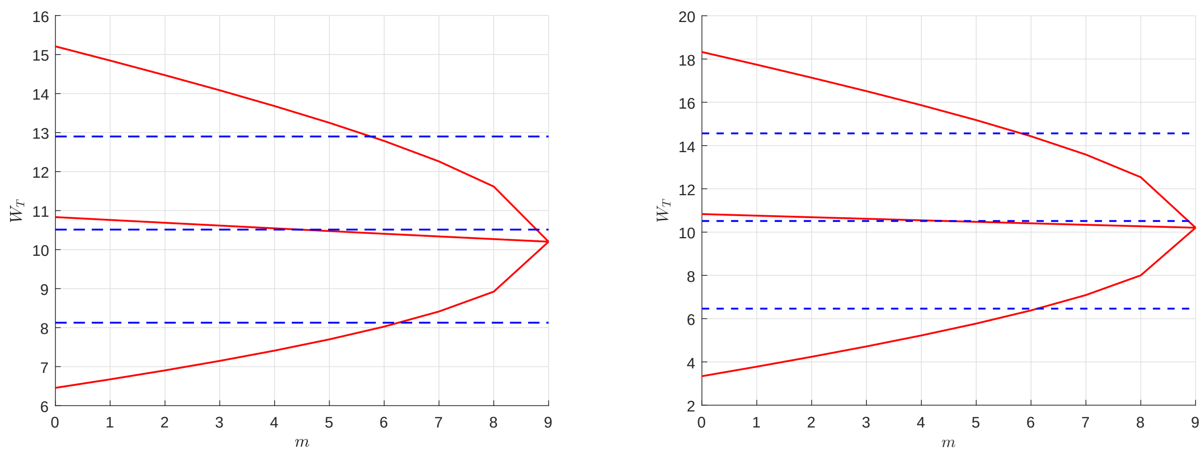

Figure 2: DCA (dashed lines) vs Lump Sum (solid lines), as a function of the period of investment, t_m, for Lump Sum. Expected terminal wealth, W_T, and two standard deviation bands. Left: GBM with $\sigma = 0.2$, $r = 0.02$, $\mu = 0.08$, $T = 1$, $W = 10$. Right: MJD with $\lambda = 2, \mu_J = -0.12, \sigma_J = 0.18$.

Corollary 2 provides a mathematical result on expected returns that is also consistent with empirical studies. In particular, it has been documented in the literature (see Williams and Bacon (1993), Knight and Mandell (1992)) that a LS investment often has higher expected return as compared to that of a DCA strategy. That is, for LS with $\alpha_0 = W$, i.e, the investor invests all of her wealth at time t_0, then from the fact that $\mathbb{E}\left[W_T^{DCA}\right]$ is decreasing in terms of M (from Corollary 2), the LS investment will have higher expect return as compared to that of a DCA investment. This higher expected return is achieved only at the cost of a higher variance of wealth, which we summarize in the following result.

Corollary 5. *(DCA vs LS) For LS_0 with $\alpha_0 = W$, we have*

$$Var(W_T^{DCA}) \leq Var(W_T^{LS_0}) = W^2\big(\exp(\psi_R(-2i)T) - \exp(2\mu T)\big). \tag{16}$$

Moreover, we can always bound the difference (uniformly in M) by

$$\mathbb{E}\left[W_T^{LS_0} - W_T^{DCA}\right] \geq \frac{W}{2}\left(e^{\mu T} - e^{rT}\right) \geq 0, \quad \forall M \geq 1. \tag{17}$$

Proof. The proof is a straightforward application of Corollary 2 and Corollary 3. □

Remark 1. *It is also interesting to note that the expected wealth for a DCA strategy is the same as that of a randomized buy and hold strategy (which is studied in Brennan et al. (2005)) which places an equal weight of investing the full amount W in any period $0 \leq m \leq M$:*

$$\mathbb{E}\left[W_T^{DCA}\right] = \sum_{m=0}^{M} \frac{1}{M+1} \mathbb{E}\left[W_T^{LS_m}\right] = \mathbb{E}[\mathbb{E}[W_T^{LS_{\mathcal{M}}}|\mathcal{M}]].$$

That is, the expected wealth of DCA is equivalent to that of a strategy which randomly chooses a period $\mathcal{M}$, and invests the full lump sum at time $\mathcal{M}$.

From Corollary 5, we see that although DCA has a lower expected return than LS, it does achieve a reduction in variance, as one would expect due to its averaging nature. This variance reduction property is attractive to many risk averse investors, and is documented in the empirical literature. The risk-return tradeoff is not so clear cut, but its analysis can be simplified with the investment strategy we introduce in section 3.3.

3.3 Geometric DCA

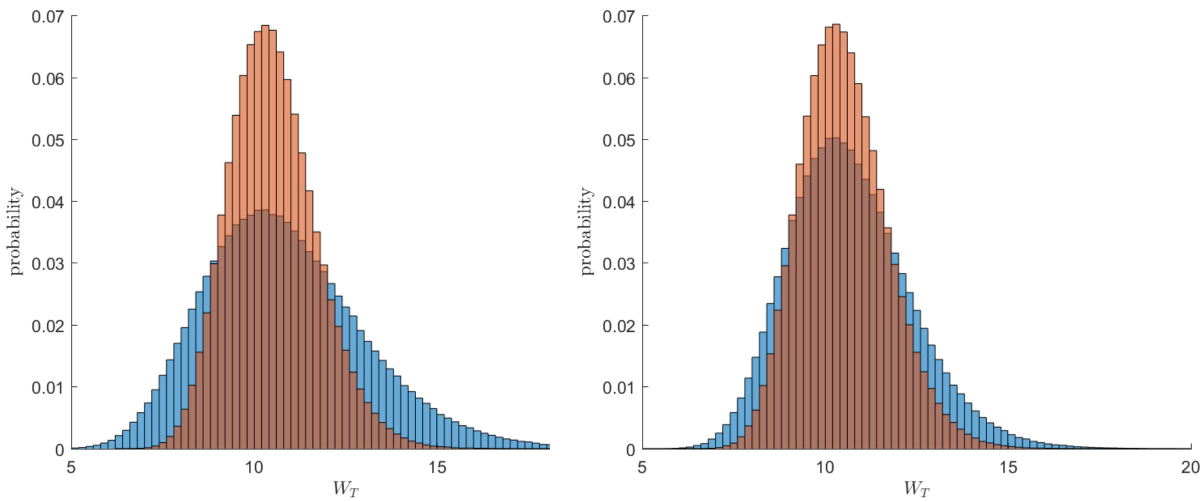

Figure 3: Distribution of terminal wealth, W_T, over 10^6 simulated paths. Model: GBM with $\sigma = 0.2$, $r = 0.02$, $\mu = 0.08$, $T = 1$, $W = 10$, $M = 9$. Left: Lump sum (blue histogram with wider variance) vs DCA (histogram with lower variance), where LS invests $\alpha_0 = W$. Right: DCA vs GDCA (histogram with lower variance) with $\theta = 0.75$.

We introduce an alternative investment strategy which highlights the nature of the relationship between DCA and LS. In particular, DCA and LS are both special cases belonging on the continuum of geometrically timed investment strategies, which we call Geometric DCA (GDCA), with weights

$$\alpha_m = W \frac{(1-\theta) \cdot \theta^m}{1 - \theta^{M+1}}, \quad m = 0, \ldots, M,$$

where $\theta \in (0,1)$ is the *duration* weight, which controls the duration of time that wealth is exposed to the risky asset over $[0,T]$. As $\theta \to 1$ (decreasing duration), we obtain DCA with $\alpha_m = W/(M+1)$, for $m = 0,\ldots,M$. As $\theta \to 0$, we obtain LS with $\alpha_0 = W$, and $\alpha_m = 0$ for $m \geq 1$. For $\theta \in (0,1)$, we obtain strategies which are a smoothly varying compromise between the duration of LS and DCA, and attain all expected returns between the two strategies. By varying the weight θ, we are able to study the impact of timing exposure to the risky asset for various risk and return measures. Moreover, GDCA might itself offer an attractive investment strategy which combines the desirable features of DCA and LS, as we will later demonstrate empirically.

Figure 3 illustrates the distribution of wealth, W_T, for each strategy. In the left figure, LS invests immediately $\alpha_0 = W$, while in the right figure GDCA is used with $\theta = 0.75$, a compromise between LS and DCA investment strategies. In either case, the distribution of DCA is much tighter around its mean of $\mathbb{E}[W_T^{DCA}] = 10.5128$, reflecting its lesser exposure (duration) to the risky asset. In the left figure, the LS strategy earns on average $\mathbb{E}[W_T^{LS}] = 10.8329$, which is about 3% in excess of the DCA strategy. To achieve this, LS is also exposed to a significantly greater level of risk, which is expected from the result obtained in Theorem 5. In the right figure, GDCA earns 1.4% in excess of DCA on average, with slightly more variance.

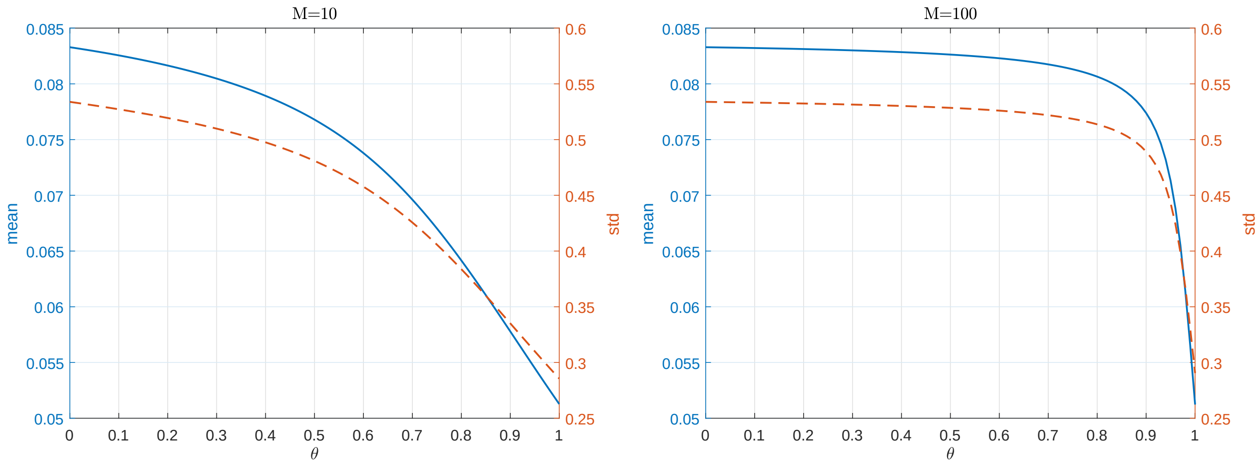

Figure 4: GDCA risk-return trade-off, as a function of θ. Mean (solid line) and std dev (dashed line) of return on initial wealth. Left: $M = 10$. Right: $M = 100$. Model: MJD with $\sigma = 0.3$, $\lambda = 3$, $\mu_J = -0.12$, $\sigma_J = 0.2$, $r = 0.02$, $\mu = 0.08$, $T = 1$, $W = 10$.

In Figure 4, we consider the effect of θ on the mean and standard deviation (std) of W_T, for two values of M. First note that for $\theta \in (0,1)$, the range of attainable values for both the mean and standard deviation of W_T are the same for $M = 10$ and $M = 100$, and as expected, both measures decline as $\theta \to 1$ (DCA). Comparing the two figures, we can see that the nature of the trade-off between risk and return as a function of θ varies greatly with M. In particular, as we approach the DCA strategy, both mean and return drop abruptly when $M = 100$, but the decline is very gradual for $M = 10$. Any measure of risk which is impacted by the two competing forces will also be greatly influenced by the choice of M.

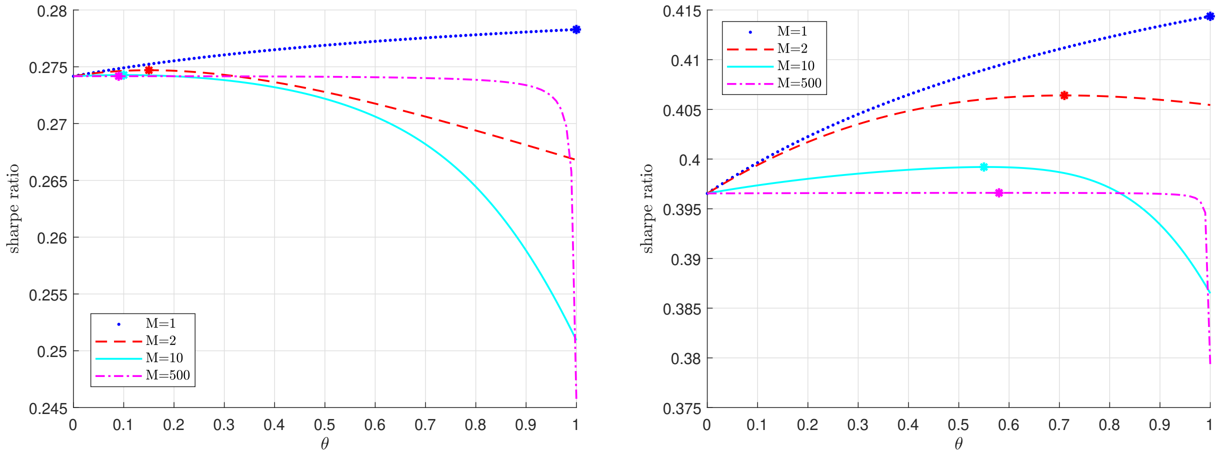

Figure 5: Sharpe ratios for GDCA as a function of θ for $M \in \{1, 2, 10, 500\}$. Left: $T = 1$. Right: $T = 3$. Model: GBM with $\sigma = 0.2$, $r = 0.02$, $\mu = 0.08$, $W = 10$.

A commonly used measure which assesses the tradeoff between excess return and risk is the Sharpe ratio, given by[8]

$$SR(\boldsymbol{\alpha}) = \frac{\log(\mathbb{E}[W_T(\boldsymbol{\alpha})]/W) - rT}{\sqrt{Var(W_T(\boldsymbol{\alpha})/W)}} \tag{18}$$

which measures the excess return minus the risk free return, per unit of wealth volatility. This metric is preferred by practitioners as it succinctly captures the trade-off in a single number, allowing investors to choose between different strategies. Figure 5 compares the Sharpe ratio for $T = 1$ (left) and $T = 3$ (right), as a function of duration θ, for $M \in \{1, 2, 10, 500\}$, where the Sharpe-optimal θ is plotted with an asterisk. As before, we note the enormous impact of M on the resulting preferred strategy. For all interior values of θ, the strategy with $M = 1$ which splits wealth between the first and last period, is preferred to other values of M for both $T = 1$ and $T = 3$. Also we observe that as $\theta \to 1$ (approaches DCA) the Sharpe ratio tends to rapidly decrease, for all $M \geq 2$.

In general, the relationship between the optimal Sharpe ratio, M, and the duration θ is complex. We note that in the examples given in Figure 5, LS is suboptimal for each value of M. With the exception of $M = 1$ (for which DCA is optimal), the optimal strategy is some compromise between DCA and LS investing, i.e. an interior $\theta^* \in (0, 1)$. This finding is later reinforced in Section 5 with an empirical study on S&P 500 returns.

[8]Note that this is a slight modification of the standard definition.

3.3.1 Effect of Interest Rates and Excess Return

This section considers the effect of interest rates and excess return on timing strategies. Given the recent state of negative interest rates in many European economies, there is an interest in understanding decision making and investing under these (formerly) unusual market conditions. Several works have explored this case in the option pricing literature, for example Battauz et al. (2018); Battauz and Rotondi (2019).

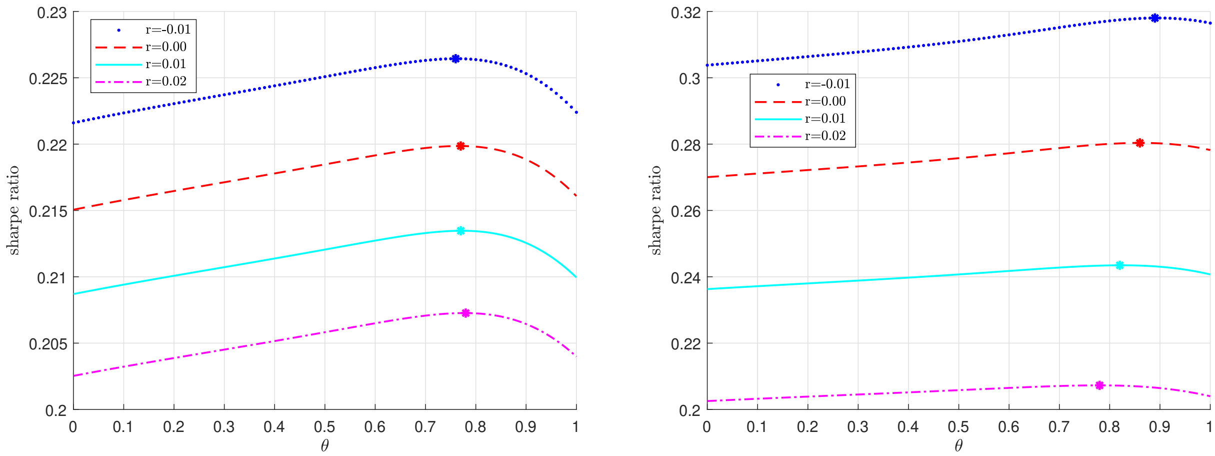

Figure 6: Sharpe ratios for GDCA as a function of θ for $r_i \in \{-0.01, 0.0, 0.01, 0.02\}$. Left: $\mu_i = 0.06 + r_i$. Right: $\mu_i = 0.08$ held fixed. Model: MJD with $\sigma = 0.3$, $\lambda = 3$, $\mu_J = -0.12$, $\sigma_J = 0.2$,$T = 3$, $W = 10$, $M = 10$.

Figure 6 illustrates the effect of rates in two interesting cases. The left panel demonstrates the effect of varying $r_i \in \{-0.01, 0.0, 0.01, 0.02\}$ when $\mu_i = 0.06 + r_i$, so the excess return is held constant ($\mu_i - r_i = 0.06$). In this case, the Sharpe-optimal θ (plotted with an asterisk) is very close for all interest rate levels, with a slight decrease in duration as rates decrease (more investment up-front). However, there is a clear change in strategy in the right panel, which corresponds to $\mu_i = 0.08$ held fixed, and the excess return is decreasing as interest rates increases. In this case, decreasing interest rates has a strong effect on shifting the optimal policy towards the DCA strategy.

3.4 Mean-Variance Optimality

We next consider the mean-variance optimal trade-off for an investor who seeks a desired (target) return, which is a classic problem stemming from Markowitz Portfolio Theory (Markowitz, 1952). Towards this end, note that we can re-express the terminal variance as

$$\begin{aligned}
Var(W_T(\boldsymbol{\alpha})) &= Var\left(\sum_{m=0}^{M}\alpha_m\frac{S_T}{S_{t_m}}\right) = Var\left(\sum_{m=0}^{M-1}\alpha_m\frac{S_T}{S_{t_m}}\right) \\
&= Var\left(\sum_{m=0}^{M-1}\alpha_m\exp(R_{m+1}+\cdots+R_M)\right) \\
&= Var\left(\sum_{m=0}^{M-1}\alpha_m X_m\right),
\end{aligned} \tag{19}$$

where $X_m := \exp(R_{m+1}+\cdots+R_M)$. Hence, we can treat each X_m as a separate "asset" in a portfolio, and these assets are correlated. The assets in this case represent a sequence of exposures to a single underlying, across overlapping segments of time. In the final time period, the return is a degenerate random variable, as any allocation of $\alpha_M > 0$ has a deterministic effect on W_T. Hence, in this section we consider a slightly modified problem in which the investor invests all wealth by period $M-1$, so that $\sum_{m=0}^{M-1}\alpha_m = W$ (rather than $\sum_{m=0}^{M}\alpha_m = W$ as we considered before). In particular, we enforce that $\alpha_M = 0$. To reformulate this as a quadratic programming problem, note that from (19) we can write

$$Var(W_T(\boldsymbol{\alpha})) = \boldsymbol{\alpha}^T H\boldsymbol{\alpha}, \tag{20}$$

where $H_{m,j} = Cov(X_m, X_j)$, which are derived in Lemma 3. Specifically, the matrix H is symmetric positive definite, with upper triangular defined as follows. For $m = 1,\ldots,M-1$,

$$H_{m,m} = \exp((M-m)\Delta t\cdot\psi_R(-2i)) - \exp(2(M-m)\Delta t\cdot\mu). \tag{21}$$

Moreover, for $1\le m\le M-1$, and $1\le k\le M-m-1$,

$$H_{m,m+k} = \exp(k\Delta t\cdot\mu)\exp((M-m-k)\Delta t\cdot\psi_R(-2i)) - \exp((2M-2m-k)\Delta t\cdot\mu). \tag{22}$$

Interpreting the exposures as correlated asset returns X_m allows us to derive the following.[9]

Theorem 6. *Let $[\widehat{\mu}_L,\widehat{\mu}_U] = [e^{rT}W, e^{\mu T}W]$ be the interval given in (9) and $\widehat{\mu}\in[\widehat{\mu}_L,\widehat{\mu}_U]$, and $\boldsymbol{\alpha} = \{\alpha_m\}_{m=0}^{M-1}\in\mathcal{F}_0$, where $M\ge 2$. Consider the following mean-variance optimization problem*[10]

$$\begin{aligned}
&\min_{\boldsymbol{\alpha}}\frac{1}{2}Var\left(W_T(\boldsymbol{\alpha})\right) \\
&\textit{Subject to}\quad \mathbb{E}[W_T(\boldsymbol{\alpha})] = \widehat{\mu},\quad \sum_{m=0}^{M-1}\alpha_m = W.
\end{aligned} \tag{23}$$

[9]Note that the optimal solution in Theorem 6 imposes no non-negativity constraints. To ensure that $\alpha_m\ge 0$, ie no short positions, we can use standard quadratic programming software to determine the optimal holding.

[10]Note that in the Theorem, $\mathbb{E}[W_T(\boldsymbol{\alpha})]$ is calculated with the formula using $M+1$ from (8), although the size of α in this optimization is only M.

Then the optimum $\boldsymbol{\alpha}^* = \{\alpha_m^*\}_{m=0}^{M-1} \in \mathcal{F}_0$ *is given by*

$$\begin{aligned}\boldsymbol{\alpha}^* = H^{-1} &\left(\frac{(\widehat{\mu} - e^{rT}W)\mathbf{1}^T H^{-1}\mathbf{1} - W\boldsymbol{y}^T H^{-1}\mathbf{1}}{(\boldsymbol{y}^T H^{-1}\boldsymbol{y})(\mathbf{1}^T H^{-1}\mathbf{1}) - (\boldsymbol{y}^T H^{-1}\mathbf{1})(\mathbf{1}^T H^{-1}\boldsymbol{y})} \right) \boldsymbol{y} \\ + H^{-1} &\left(\frac{W\boldsymbol{y}^T H^{-1}\boldsymbol{y} - (\widehat{\mu} - e^{rT}W)\mathbf{1}^T H^{-1}\boldsymbol{y}}{(\boldsymbol{y}^T H^{-1}\boldsymbol{y})(\mathbf{1}^T H^{-1}\mathbf{1}) - (\boldsymbol{y}^T H^{-1}\mathbf{1})(\mathbf{1}^T H^{-1}\boldsymbol{y})} \right) \mathbf{1}, \end{aligned} \tag{24}$$

where $\boldsymbol{y} = \left(e^{\mu T} - e^{rT}, e^{\mu(T-t_1)} - e^{r(T-t_1)}, \ldots, e^{\mu(T-t_{M-1})} - e^{r(T-t_{M-1})}\right)^T$.

Proof. See appendix B. □

Remark 2. *Note that in Theorem 6, aside from* $\alpha_M = 0$ *and* $\sum_{m=0}^{M-1} \alpha_m = W$*, we enforce no sign constraints on* $\boldsymbol{\alpha}$*, which permits short positions to be taken. However,* (23) *can be solved with standard quadratic programming software if we incorporate the constraints* $\alpha_m \geq 0$*,* $m = 0, \ldots, M-1$*.*

To illustrate the applicability of Theorem 6, consider an investor who targets a terminal return of $\widehat{\mu} \in [\widehat{\mu}_L, \widehat{\mu}_U]$, which can be parameterized by $\zeta \in (0,1)$ as the convex combination

$$\widehat{\mu}(\zeta) = \zeta\widehat{\mu}_L + (1-\zeta)\widehat{\mu}_U = W(\zeta \exp(rT) + (1-\zeta)\exp(\mu T)). \tag{25}$$

Consider the case of $W = 1$, so that α_m is the proportion of the initial wealth invested at t_m. As an example, we consider an investor who invests in quarterly installments over $T = 3$ years, so $M = 12$. Figure 7 illustrates two preferences, with $\zeta = 0.5$ in the left figure, and $\zeta = 0.75$ on the right. In either case, the optimal strategy involves an investment in each period, with the boundary periods dominating the overall allocation.

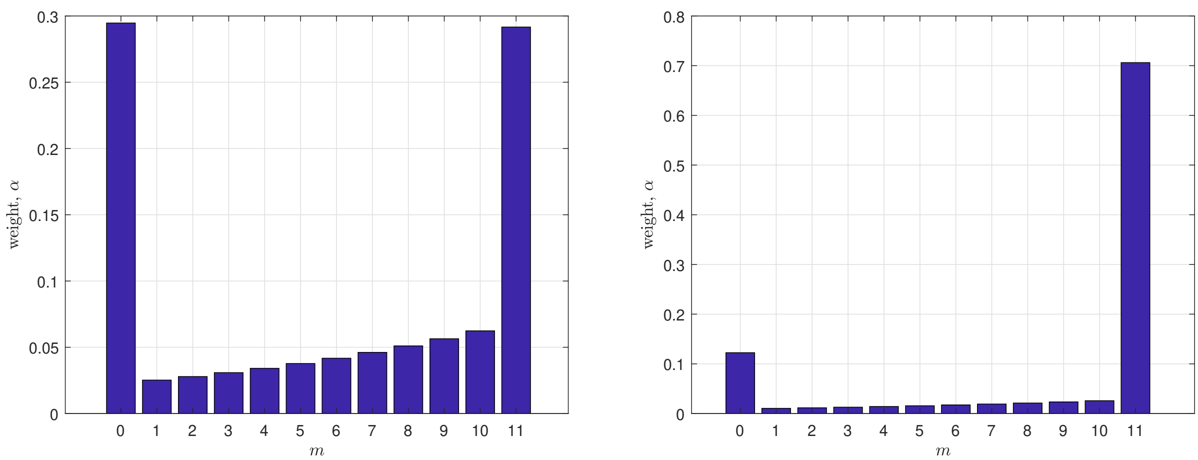

Figure 7: Mean-variance optimal weights $\boldsymbol{\alpha}$. Parameters: $W = 1$, $r = 0.02$, $\mu = 0.08$, $M = 12$ and $T = 3$. Target return $\widehat{\mu}(\zeta)$ defined in (25). Left: $\zeta = 0.5$. Right: $\zeta = 0.75$. Model: MJD with $\sigma = 0.3$, $\lambda = 2$, $\mu_J = -0.12$, $\sigma_J = 0.3$.

3.5 Equivalence to Arithmetic Averaging

In this section, we establish a fundamental distributional equivalence between the stochastic component W_T^S and arithmetic averaging. This equivalence not only simplifies the computation of risk measures, but also opens the door for hedging strategies based on (liquidly) traded arithmetic Asian options. We assume that potential investments are made at regularly spaced time intervals, so that $t_{m+1} - t_m = \Delta t$, for $m = 0, \ldots, M-1$, for example once per day.

Theorem 7. *Let $\Delta t := t_{m+1} - t_m$ be uniform. Then under exponential Lévy dynamics,*

1. *For $\boldsymbol{\alpha} \in \mathcal{F}_0$, the equivalence holds:*

$$\sum_{m=0}^{M} \alpha_{M-m} S_{t_m} \stackrel{d}{=} S_0 \cdot \sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}} \tag{26}$$

2. *Define the averages*

$$A_T := \frac{1}{M+1} \sum_{m=0}^{M} S_{t_m}, \quad H_T := \frac{1}{M+1} \sum_{m=0}^{M} \frac{S_T}{S_{t_m}}.$$

Let $\mathcal{V}_A(S_0, K, M, T)$ and $\mathcal{V}_H(S_0, L, M, T)$ denote the value of an Asian option written on A_T and H_T, respectively. Then we have the pricing equation

$$\mathcal{V}_H(S_0, L, M, T) = \frac{e^{-rT}}{S_0} \cdot \mathcal{V}_A(S_0, S_0 \cdot L, M, T). \tag{27}$$

Proof. See appendix B. □

In particular, we can hedge the risk of the DCA strategy with an initial wealth of W, and $\alpha_m = W/(M+1)$, using

$$\left(\sum_{m=0}^{M} \frac{W}{M+1} \frac{S_T}{S_{t_m}} - L \right)^+ = W \left(\sum_{m=0}^{M} \frac{S_T}{S_{t_m}} - \frac{L}{W} \right)^+$$

As Asian options are liquid in many traded markets, especially commodities and energy markets, this implies that the risk of a DCA trading strategy may be offset (hedged) by positions in Asian options.[11]

4 The PROJ Method for Timing Strategy Risk

In this section, we apply the frame projection (PROJ) method introduced in Kirkby (2015); Kirkby and Deng (2019) to estimate the risk of DCA and other deterministic timing strategies.

[11] We note that the PROJ method presented below can also be used to value options written on DCA. In this case, one must switch to the risk-neutral measure, which for the proposed exponential Lévy model is a simple as changing the drift from μ to r.

It also allows for the determination of optimal $\boldsymbol{\alpha} \in \mathcal{F}_0$, based on commonly used risk measures such as Value-at-Risk and Downside Risk. The PROJ method has been successfully applied in several contexts involving exotic option pricing Cui et al. (2019b, 2017); Kirkby (2018), actuarial theory Wang and Zhang (2019); Zhang et al. (2020), as well as non-parametric density estimation Cui et al. (2019c). In the following section, we briefly review the PROJ method, and then describe how it can be used to recover the distribution of $H_T(\boldsymbol{\alpha}) := \sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}}$. This will allow us to calculate arbitrary risk measures, such as value at risk and expected shortfall, for any deterministic timing strategy with weights $\boldsymbol{\alpha} \in \mathcal{F}_0$.

4.1 Frame Projection

Given a random variable Y whose (unknown) probability density function is denoted by $f_Y \in L^2(\mathbb{R})$ but its ChF function $\phi_Y(\xi) = \mathbb{E}[e^{i\xi Y}]$ is known or can be approximated in some form. Let φ be a compactly supported function, which is often called a generator. For example, φ can be chosen as linear spline function as below

$$\varphi(y) = (1+y)\mathbb{1}_{[-1,0]}(y) + (1-y)\mathbb{1}_{[0,1]}(y).$$

Choose a *resolution* $a > 0$, and a reference point $-\infty < y_1 < \infty$. Using a, y_1 and φ, we generate a set of compactly supported functions $\{\varphi_{a,n}\}_{n\in\mathbb{Z}}$ obtained from the generator φ using dilation and translation transformations, $\varphi_{a,n}(y) := a^{1/2}\varphi(a(y-y_n))$, where $y_n = y_1 + (n-1)/a$ for $n \in \mathbb{Z}$. Now consider the subspace $\mathcal{M}_a := \overline{\text{span}}\{\varphi_{a,n}\}_{n\in\mathbb{Z}}$ of $L^2(\mathbb{R})$. Given $f_Y \in L^2(\mathbb{R})$, the orthogonal projection of f_Y, denoted by $P_{\mathcal{M}_a} f_Y$, onto $\mathcal{M}_a$ is determined by the *dual basis* $\{\widetilde{\varphi}_{a,n}\}_{n\in\mathbb{Z}}$:

$$P_{\mathcal{M}_a} f_Y(y) = \sum_{n\in\mathbb{Z}} \langle f_Y, \widetilde{\varphi}_{a,n}\rangle \varphi_{a,n}(y). \tag{28}$$

In (28), the dual basis $\{\widetilde{\varphi}_{a,n}\}_{n\in\mathbb{Z}}$ is *biorthogonal* in the sense that $\langle \varphi_{a,n}, \widetilde{\varphi}_{a,m}\rangle = \mathbb{1}_{\{n=m\}}$, as well as the particular case of an orthogonal basis is self-dual. Moreover, the projection coefficients $\langle f_Y, \widetilde{\varphi}_{a,n}\rangle$'s in (28) are derived in closed-form using the Fourier transform $\widehat{\widetilde{\varphi}}$ of $\widetilde{\varphi}$:

$$\langle f_Y, \widetilde{\varphi}_{a,n}\rangle = \frac{a^{-1/2}}{\pi} \Re \left[\int_0^\infty \exp(-\mathrm{i} y_n \xi) \cdot \phi_Y(\xi) \widehat{\widetilde{\varphi}}\Big(\frac{\xi}{a}\Big) d\xi \right], \tag{29}$$

given that $\widehat{\widetilde{\varphi}}(\xi)$ is known. For example, for the linear splines, we have

$$\widehat{\widetilde{\varphi}}(\xi) = 12\sin^2(\xi/2)/(\xi^2(2+\cos(\xi))). \tag{30}$$

Next, choosing $N \in \mathbb{N}^+$, given a and y_1 have been chosen appropriately, we confine $\mathcal{M}_a$ to a domain of finite set $\{y_n\}_{n=1}^{N}$ where each basis function $\varphi_{a,n}(y)$ is centered around the grid point $y_n = y_1 + (n-1)/a$. In addition, to account for the Nyquist frequency, an N-point frequency grid is specified by $\Delta_\xi = 2\pi a/N, \xi_n = (n-1)\Delta_\xi, n = 1, ..., N$, and this is numerically applied to invert the analytical coefficient representation in equation (29). Therefore the (unknown) probability density function f_Y of the random variable Y can be now approximated by

$$f_Y(y) \approx P_{\mathcal{M}_a} f_Y(y) \approx \sum_{1\leq n\leq N} \bar{\beta}_{a,n}\varphi_{a,n}(y). \tag{31}$$

With constant $\Upsilon_{a,N} := 24a^2/N$, the coefficients $\bar{\beta}_{a,n} \approx \langle f_Y, \widetilde{\varphi}_{a,n}\rangle$ are determined as follows

$$\left\{\bar{\beta}_{a,n}\right\}_{n=1}^{N} := a^{1/2}\Upsilon_{a,N}\Re\left\{\mathcal{D}\left\{G_j\right\}\right\}, \quad \mathcal{D}_n\left\{G_j\right\} = \sum_{j=1}^{N} e^{-\mathrm{i}\frac{2\pi}{N}(j-1)(n-1)}G_j, \quad n = 1, ..., N. \tag{32}$$

In (32), $\mathcal{D}\{.\}$ denotes the discrete Fourier transform (DFT) operator. The input vectors $\{G_j\}_{j=1}^N$ is specified by

$$G_1 := 1/24a^2, \qquad G_m := \phi_Y(\xi_n)\cdot\mathcal{B}_n, \quad n \geq 2, \tag{33}$$

where

$$\mathcal{B}_n := \frac{(\sin(\xi_n/(2a))/\xi_n)^2}{2+\cos(\xi_n/a)}, \quad n \geq 2. \tag{34}$$

For details on the derivation of $\widehat{\widetilde{\varphi}}$ (and $\mathcal{B}_n$) for B-spline bases, the reader is invited to refer to Kirkby (2017).

4.2 Characteristic Function Recovery

We now briefly describe the procedure required to calculate Y_M using the PROJ method. More details can be found in Kirkby (2016), Kirkby and Nguyen (2020), where the current calculation is closely related to that of an Asian option. This recursive approach can also be applied with alternative Fourier techniques, see for example Leitao et al. (2019) for a related application.

The next result enables us to easily calculate the distribution of

$$H_T(\boldsymbol{\alpha}) := \sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}} \tag{35}$$

which in turn enables the computation of arbitrary risk measures on W_T, since

$$W_T(\boldsymbol{\alpha}) = H_T(\boldsymbol{\alpha}) + e^{rT}W - \sum_{m=0}^{M}\alpha_m e^{r(T-t_m)} := H_T(\boldsymbol{\alpha}) + W^R(\boldsymbol{\alpha}), \tag{36}$$

where $W^R(\boldsymbol{\alpha})$ is deterministic.

Corollary 8. *Suppose that $\alpha_m > 0$ for $m = 0, \ldots, M$. Set $Y_1 = \log(\alpha_0/\alpha_1) + R_M$, and define recursively*

$$Y_m = \log\left(\frac{\alpha_{m-1}}{\alpha_m}\right) + R_{M+1-m} + Z_{m-1}, \quad m = 2, \ldots, M, \tag{37}$$

where $R_m = \log(S_{t_m}/S_{t_{m-1}})$ and $Z_m := \log(1+\exp(Y_m))$. Then

$$H_T(\boldsymbol{\alpha}) \stackrel{d}{=} \alpha_M\left(1+\exp(Y_M)\right). \tag{38}$$

Proof. See appendix B. □

From Corollary 8, we obtain a recursive formula for the ChF of Y_m defined in (37), starting with Y_1. It is easy to see the ChF of Y_1 is given by

$$\phi_{Y_1}(\xi) = e^{i\xi \log\left(\frac{\alpha_0}{\alpha_1}\right)} e^{i\xi R_M} = e^{i\xi \log\left(\frac{\alpha_0}{\alpha_1}\right)} \phi_R(\xi),$$

which initializes the recursion. Similarly, applying (37) we obtain the formula

$$\phi_{Y_m}(\xi) = \phi_R(\xi)\phi_{Z_{m-1}}(\xi) e^{i\xi \log(\frac{\alpha_{m-1}}{\alpha_m})}, \quad m = 2, \dots, M. \tag{39}$$

In the above, the ChF $\phi_{Z_{m-1}}(\xi)$ of Z_{m-1} is given by

$$\phi_{Z_m}(\xi) = \mathbb{E}[e^{i\xi \log(1+\exp(Y_m))}] = \int (e^y + 1)^{i\xi} f_{Y_m}(y) dy, \tag{40}$$

which can be calculated in closed form as described in Kirkby (2016).

4.3 Risk Measurement

We now describe how to use the PROJ method for the calculation of various risk measures, such as those satisfying the coherency axioms formalized in (Artzner et al., 2003). Risk measures are commonly used to identify an optimal portfolio strategy, such as Value-at-Risk Staino and Russo (2019), Downside Risk Ling et al. (2019), and mean-variance criteria Lwin et al. (2017); Bi et al. (2018); Penev et al. (2019). Given a risk measure $\rho : \mathbb{R} \to \mathbb{R}$, we can calculate

$$\rho\left(W_T(\boldsymbol{\alpha})\right) = \rho\left(H_T(\boldsymbol{\alpha}) + W^R(\boldsymbol{\alpha})\right) = \rho\left(\sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}} + W^R(\boldsymbol{\alpha})\right), \tag{41}$$

where $W^R(\boldsymbol{\alpha})$ is deterministic, which follows from (36). Note that any translation invariant risk measure ρ (for example a coherent measure, (Artzner et al., 2003; Brandtner et al., 2018)), we have that

$$\rho\left(W_T(\boldsymbol{\alpha})\right) = \rho\left(\sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}}\right) + W^R(\boldsymbol{\alpha}),$$

since $W^R(\boldsymbol{\alpha})$ is non-stochastic. However, more generally we will use (41), as it allows us to compute a wider variety of measures.

Utilizing the the relation from Corollary 8

$$H_T(\boldsymbol{\alpha}) = \sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}} \stackrel{d}{=} \alpha_M \left(1 + \exp(Y_M)\right),$$

and applying the recursive procedure described in Section 4.2, we obtain the cf of Y_M, $\phi_{Y_M}(\xi)$ from (39). We can then use (31), with $\phi_{Y_M}(\xi)$ in place of $\phi_Y(\xi)$, to obtain the projected density with coefficients of Y_M, $\{\bar{\beta}_{a,n}\}_{n=1}^N$.

We illustrate the calculation of risk for Value-at-Risk in the next section. Other risk measures, such as moments and expected shortfall (see (Szegö, 2005)), are similarly easy to calculate using this approach, and also result in closed form approximations.

4.3.1 Value-at-Risk

Value-at-Risk (VaR) is a conventional risk measure to assess the riskiness of a portfolio, and is commonly used as a tool for portfolio selection and asset allocation Babat et al. (2018); Staino and Russo (2019); Ahmadi-Javid and Fallah-Tafti (2019); Cui et al. (2019a). To calculate VaR, defined by

$$VaR_\rho(H_T(\boldsymbol{\alpha})) = \inf\{x \in \mathbb{R} : \mathbb{P}(H_T(\boldsymbol{\alpha}) \le x) \ge \rho\}, \tag{42}$$

we take advantage of one of the nice properties of the linear basis. In particular, define the cumulative distribution approximation of Y_M, which is calculated at the grid points $\{y_n\}_{n=1}^N$ using

$$\bar{F}_n := \bar{F}(y_n) = a^{-1/2} \sum_{j=1}^{n-1} \bar{\beta}_{a,j} + \frac{a^{-1/2}}{2} \bar{\beta}_{a,n}, \tag{43}$$

and define the boundary coefficients $\bar{\beta}_{a,0} = \bar{\beta}_{a,N+1} = 0$. We can then calculate the distribution at any $y \in \mathbb{R}$ using

$$\bar{F}(y) = \bar{F}(y_n) + a^{-1/2} \left[\gamma \bar{\beta}_{a,n} + \frac{\gamma^2}{2} \left(\bar{\beta}_{a,n+1} - \bar{\beta}_{a,n} \right) \right],$$

$\gamma := a(y - y_n)$, and is defined as $F(y) = 1$ for $y > y_N$ and $F(y) = 0$ for $y < y_1$. For any VaR level, $\rho \in (0, 1)$, let $k \in \{0 \dots, N\}$ be the unique integer satisfying $\bar{F}_k \le \rho < \bar{F}_{k+1}$, and set $d_k := \bar{\beta}_{a,k+1} - \bar{\beta}_{a,k}$. We then have closed-form expression for VaR (see Cui et al. (2018)) in the case of the linear basis,

$$VaR_\rho(Y_M) = \begin{cases} y_k + \dfrac{1}{a \cdot d_k} \left(-\bar{\beta}_{a,k} + \sqrt{\bar{\beta}_{a,k}^2 + 2a^{1/2} \cdot d_k(\rho - \bar{F}_k)} \right), & d_k \neq 0; \\ y_k + \dfrac{\rho - \bar{F}_k}{a \cdot (\bar{F}_{k+1} - \bar{F}_k)}, & d_k = 0. \end{cases} \tag{44}$$

From the identity $H_T(\boldsymbol{\alpha}) \overset{d}{=} \alpha_M (1 + \exp(Y_M))$, and the fact that monotone transformations preserve quantiles,

$$VaR_\rho(H_T(\boldsymbol{\alpha})) = \alpha_M(1 + \exp(VaR_\rho(Y_M)))$$

where for DCA we have $\alpha_M = W/(M + 1)$. Similarly, we can calculate $VaR_\rho(W_T(\boldsymbol{\alpha})) = VaR_\rho(H_T(\boldsymbol{\alpha})) + W^R(\boldsymbol{\alpha})$.

To illustrate the recursive PROJ approach for calculating risk, we provide an experiment for GBM which assess VaR. In Figure 8, we see another manifestation of the risk reduction of DCA, which is a tighter terminal wealth distribution. The tightness varies smoothly as a function of θ, which we plot for $M = 10$ and $M = 100$.

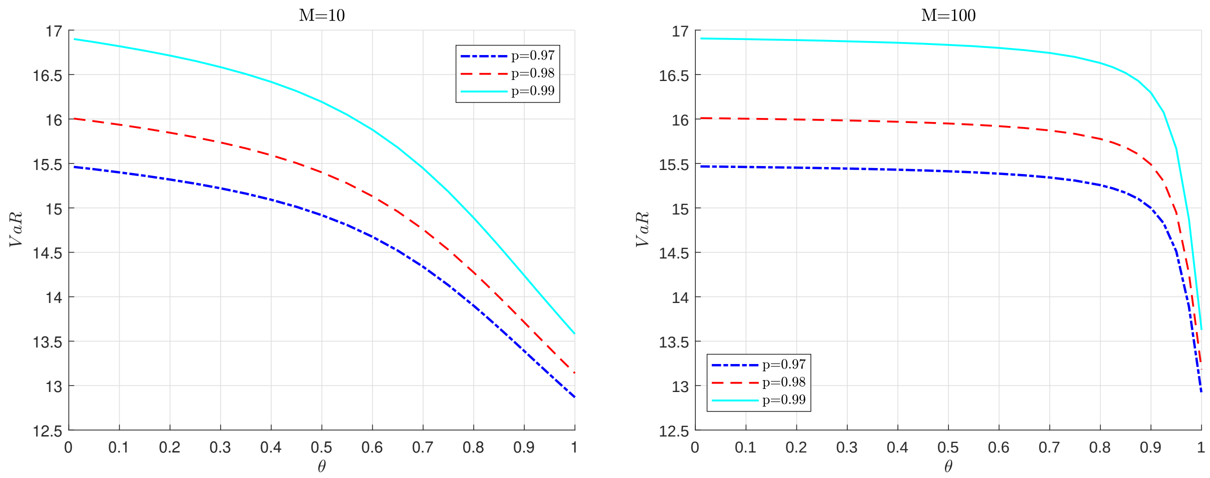

Figure 8: GDCA value at risk (VaR), as a function of θ, for $\rho \in \{0.97, 0.98, 0.99\}$. Left: $M = 10$. Right: $M = 100$. Model: GBM with $\sigma = 0.2$, $\mu = 0.08$, $T = 1$, $W = 10$.

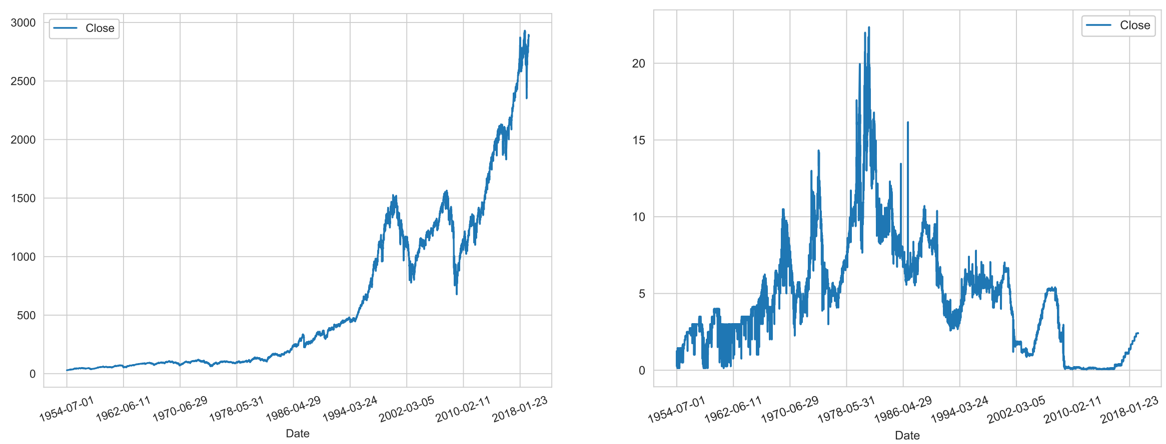

Figure 9: S&P500 index (Left) and fed funds rate (Right): 1954/07/01 - 2019/04/11.

5 Empirical Study of Investment Strategies

We now conduct an empirical study to assess the performance of DCA, GDCA, and LS investment strategies. For the risky asset, we take the S&P500 index (SPX), and for the risk-free rate we use the federal funds rate, for which the targeted rate is set by the Federal Open Market committee (FOMC). Both series are observed over the period 1954-07-01 to 2019-04-11, for a total of 16,305 observations, and both datasets are plotted in Figure 9. We assume that investments in the risk-free account accrue interest at the prevailing fed funds rate over the next day. For convenience, we introduce the notation $\bar{M}$, to denote the number of investments *per year* (in addition to the initial investment), so that $M = 1 + \bar{M} \cdot T$ (which accounts for the possible investment at t_0). Hence, monthly investing $\bar{M} = 12$ results in $M = 13$ investments over one year ($T = 1$), $M = 25$ investments over two years ($T = 2$), and so on.[12]

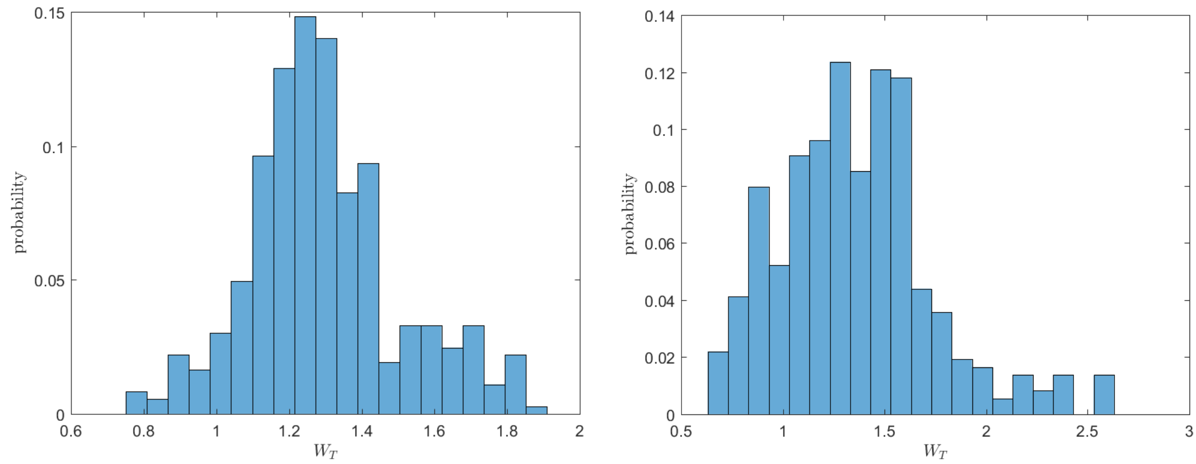

Figure 10: Terminal wealth distributions for S&P500 example, $T = 4$, $W = 1$. (Left): DCA(1), (Right): Lump Sum (LS).

For each investment horizon T, we form an empirical distribution of terminal wealth by sliding the investment window forward in time, two months at a time, which results in $N = 382$ trials.[13] This has the effect of randomizing the starting and ending periods. We denote the wealth by $W_{i,T}$, for trials $i = 1, \ldots, N$. In addition to the Sharpe ratio, mean, and variance of $\{W_{i,T}\}_{i=1}^{N}$, we introduce the power utility

$$u(W_T) = \frac{W_T^{1-\gamma}}{1-\gamma}, \tag{45}$$

which is considered for example in Brennan et al. (2005). Here γ is the coefficient of relative risk aversion. The investor becomes more risk averse for larger γ. We then define the *certainty equivalent*

$$CE_\gamma := \left(\frac{1}{N} \sum_{i=1}^{N} W_{i,T}^{1-\gamma} \right)^{\frac{1}{1-\gamma}}$$

where $W_{i,T}$ is the terminal wealth calculated for trial i at horizon T, and N is the total number of trials. The certainty equivalent for a given strategy is the amount of wealth that the investor would have to receive with certainty at T in order to be indifferent between the certain payoff and the investment strategy.

We first compare the lump sum (LS) strategy and DCA for three levels of $\bar{M}$, which we denote by DCA($\bar{M}$). For example, DCA(1) is a strategy which involves an initial investment at t_0, as well as an investment at the end of each year.[14] As a benchmark, we provide a pure risk-free investing strategy (RF), where the investor earns interest at the risk-free rate which

[12] For simplicity, we assume that there are 252 trading days in each year, and 21 days in each month.
[13] Note that these trials have overlapping investment periods, and are thus not an iid random sample.
[14] Note that DCA(1) is the 50:50 strategy when $T = 1$.

resets daily. All strategies start with an initial wealth of $W_0 = W = 1$, so the terminal wealth is the same as the return. Figure 10 displays the terminal wealth distributions for DCA(1) (Left) and LS (Right), for $T = 4$. As expected, the distribution of LS is much wider than that of DCA(1), offering the potential for greater returns (and losses). The fatter left tail of LS also implies that the LS is a riskier strategy compared to DCA(1).

Horizon	strategy	mean	std	Sharpe	CE_2	CE_4	CE_6
$T = 1$	RF	1.051	0.038	0.000 (0.000)	1.049 (0.002)	1.048 (0.002)	1.047 (0.002)
(12 months)	LS	**1.084**	0.161	**0.200 (0.055)**	1.058 (0.009)	1.026 (0.012)	0.988 (0.015)
	DCA(1)	1.068	**0.081**	**0.200 (0.054)**	**1.061 (0.004)**	**1.054 (0.005)**	**1.047 (0.005)**
	DCA(6)	1.067	0.088	0.180 (0.054)	1.059 (0.005)	1.051 (0.006)	1.040 (0.007)
	DCA(12)	1.067	0.089	0.182 (0.055)	1.059 (0.005)	1.050 (0.006)	1.040 (0.007)
$T = 2$	RF	1.106	0.078	0.000 (0.000)	1.101 (0.004)	1.096 (0.004)	1.092 (0.004)
(24 months)	LS	**1.169**	0.232	**0.254 (0.054)**	1.117 (0.014)	1.053 (0.018)	0.978 (0.023)
	DCA(1)	1.137	**0.123**	0.246 (0.056)	**1.123 (0.007)**	**1.107 (0.007)**	**1.090 (0.009)**
	DCA(6)	1.137	0.131	0.229 (0.056)	1.120 (0.007)	1.101 (0.009)	1.079 (0.011)
	DCA(12)	1.137	0.132	0.230 (0.057)	1.120 (0.007)	1.101 (0.009)	1.078 (0.012)
$T = 3$	RF	1.167	0.120	0.000 (0.000)	1.156 (0.006)	1.145 (0.005)	1.136 (0.005)
(36 months)	LS	**1.257**	0.300	**0.280 (0.051)**	1.183 (0.016)	1.101 (0.019)	1.019 (0.021)
	DCA(1)	1.212	**0.165**	0.275 (0.055)	**1.189 (0.009)**	**1.164 (0.010)**	**1.138 (0.011)**
	DCA(6)	1.211	0.170	0.265 (0.055)	1.186 (0.009)	1.158 (0.011)	1.127 (0.014)
	DCA(12)	1.211	0.171	0.265 (0.057)	1.186 (0.009)	1.158 (0.011)	1.125 (0.014)
$T = 4$	RF	1.233	0.165	0.000 (0.000)	1.213 (0.008)	1.195 (0.007)	1.179 (0.007)
(48 months)	LS	**1.356**	0.382	0.300 (0.050)	1.254 (0.019)	1.159 (0.020)	1.077 (0.019)
	DCA(1)	1.294	**0.209**	**0.302 (0.051)**	**1.261 (0.011)**	**1.228 (0.012)**	**1.193 (0.014)**
	DCA(6)	1.294	0.213	0.295 (0.054)	1.259 (0.011)	1.222 (0.012)	1.182 (0.016)
	DCA(12)	1.294	0.213	0.296 (0.053)	1.259 (0.011)	1.221 (0.013)	1.181 (0.016)

Table 1: Investment strategies for S&P500 example, with initial wealth $W = 1$.

In Table 1, for each of the horizons $T \in \{1, 2, 3, 4\}$, we display the top strategy in bold for each metric (excluding RF, which is simply a benchmark). In addition to each metric, we display a standard error (in parenthesis), estimated by non-parametric bootstrap from the terminal wealth distributions. As expected from (17), as well as existing literature such as Williams and Bacon (1993), LS results in the greatest mean return for each horizon. However, this greater return is realized at the cost of greater variance, which is consistent with Theorem 5. The tradeoff, as measured by the Sharpe ratio[15], favors LS and DCA(1), depending on the horizon. However, with respect to the power utility certainty equivalents with $\gamma \in \{2, 4, 6\}$,

[15]The Sharpe ratio is computed with respect to the risk-free (RF) investment strategy. As the RF strategy has zero excess return against itself, it has a Sharpe ratio of zero.

DCA(1) is clearly favorable for every horizon. The margin by which DCA(1) is favored over LS increases with γ, that is as an investor becomes more risk averse the greater the favorability of DCA(1) increases. While the mean is very similar for the DCA strategies, we are able to confirm the theoretical result of Corollary 2 that the mean decreases as M (or $\bar{M}$) increases. The tendency for variance to increase (as illustrated in Figure 1) is also reflected in this data set. We can also see that the mean and variance of the DCA strategies (quickly) approach a limit as M increases, consistent with Corollary 2 and Corollary 4.

Horizon	strategy	mean	std	Sharpe	CE_2	CE_4	CE_6
$T = 1$	DCA(1)	1.068	**0.081**	**0.200 (0.054)**	**1.061 (0.004)**	**1.054 (0.005)**	**1.047 (0.005)**
(12 months)	LS	**1.084**	0.161	**0.200 (0.055)**	1.058 (0.009)	1.026 (0.011)	0.988 (0.015)
	$\text{GDCA}_{0.25}(1)$	1.078	0.129	**0.200 (0.054)**	**1.061 (0.007)**	1.043 (0.008)	1.021 (0.010)
	$\text{GDCA}_{0.25}(12)$	1.084	0.156	**0.201 (0.054)**	1.058 (0.009)	1.028 (0.011)	0.992 (0.015)
	$\text{GDCA}_{0.50}(12)$	1.082	0.148	**0.200 (0.054)**	1.059 (0.009)	1.032 (0.010)	0.999 (0.014)
	$\text{GDCA}_{0.75}(12)$	1.077	0.127	0.195 (0.055)	**1.060 (0.007)**	1.040 (0.009)	1.016 (0.012)
	DCA(12)	1.067	0.089	0.182 (0.056)	1.059 (0.005)	1.050 (0.006)	1.040 (0.007)
$T = 2$	DCA(1)	1.137	**0.123**	0.246 (0.055)	**1.123 (0.007)**	**1.107 (0.007)**	**1.090 (0.009)**
(24 months)	LS	**1.169**	0.232	**0.254 (0.054)**	1.117 (0.014)	1.053 (0.017)	0.978 (0.023)
	$\text{GDCA}_{0.25}(1)$	1.160	0.197	**0.255 (0.054)**	**1.122 (0.012)**	1.077 (0.015)	1.022 (0.019)
	$\text{GDCA}_{0.25}(12)$	**1.168**	0.229	**0.255 (0.054)**	1.118 (0.014)	1.055 (0.018)	0.981 (0.023)
	$\text{GDCA}_{0.50}(12)$	1.167	0.223	**0.254 (0.054)**	1.118 (0.013)	1.058 (0.017)	0.987 (0.022)
	$\text{GDCA}_{0.75}(12)$	1.161	0.205	0.251 (0.055)	1.120 (0.012)	1.068 (0.016)	1.005 (0.021)
	DCA(12)	1.137	0.132	0.230 (0.058)	1.120 (0.008)	1.101 (0.009)	1.078 (0.012)
$T = 3$	DCA(1)	1.212	**0.165**	0.275 (0.054)	**1.189 (0.009)**	**1.164 (0.010)**	**1.138 (0.011)**
(36 months)	LS	**1.257**	0.300	**0.280 (0.051)**	1.183 (0.016)	1.101 (0.020)	1.019 (0.021)
	$\text{GDCA}_{0.25}(1)$	1.247	0.264	**0.283 (0.052)**	**1.188 (0.015)**	1.122 (0.018)	1.050 (0.021)
	$\text{GDCA}_{0.25}(12)$	**1.256**	0.296	**0.281 (0.052)**	1.183 (0.016)	1.104 (0.019)	1.022 (0.022)
	$\text{GDCA}_{0.50}(12)$	1.254	0.289	**0.281 (0.052)**	1.184 (0.016)	1.106 (0.019)	1.026 (0.021)
	$\text{GDCA}_{0.75}(12)$	1.249	0.271	**0.281 (0.052)**	1.186 (0.015)	1.114 (0.019)	1.037 (0.021)
	DCA(12)	1.211	0.171	0.265 (0.056)	1.186 (0.010)	1.158 (0.011)	1.125 (0.014)
$T = 4$	DCA(1)	1.294	**0.209**	**0.302 (0.054)**	**1.261 (0.011)**	**1.228 (0.011)**	**1.193 (0.014)**
(48 months)	LS	**1.356**	0.382	0.299 (0.050)	1.254 (0.019)	1.159 (0.019)	1.077 (0.019)
	$\text{GDCA}_{0.25}(1)$	1.345	0.344	**0.304 (0.050)**	**1.261 (0.017)**	1.178 (0.019)	1.101 (0.020)
	$\text{GDCA}_{0.25}(12)$	**1.355**	0.378	**0.301 (0.049)**	1.255 (0.018)	1.162 (0.019)	1.080 (0.019)
	$\text{GDCA}_{0.50}(12)$	1.353	0.371	**0.301 (0.049)**	1.256 (0.019)	1.164 (0.019)	1.083 (0.020)
	$\text{GDCA}_{0.75}(12)$	1.347	0.352	**0.301 (0.049)**	1.258 (0.018)	1.171 (0.019)	1.091 (0.020)
	DCA(12)	1.294	0.213	0.296 (0.054)	1.259 (0.011)	1.221 (0.013)	1.181 (0.016)

Table 2: GCDA investment strategies for S&P500 example, with initial wealth $W = 1$.

In the second set of experiments, we consider the GDCA strategy, which we also compare to

DCA(1), DCA(12), and LS to provide a comparison of performance. The notation $\text{GDCA}_\theta(\bar{M})$ denotes a strategy with duration parameter θ and investments per year $\bar{M}$. As before, LS has the greatest mean return in most cases, but $\text{GDCA}_{0.25}(1)$ is often a close second, and often outperforms based on the Sharpe ratio, which indicates a more "efficient" exposure to risk. The certainy equivalence measures still favor DCA(1), but in several cases $\text{GDCA}_{0.25}(1)$ is a dcomparable strategy with respect to CE_2, i.e. with moderate risk aversion. Consequently, one may conclude that, depending on our choice of risk measurement, a strategy in between LS and standard DCA may offer the best trade-off in terms of risk and return.

6 Conclusion

In this paper, we develop a rigorous and theoretical framework for the analysis of DCA, LS, and other related investment timing strategies. We present new theoretical insights into the nature of DCA and averaging-style investment timing strategies. We examine them under non-trivial and realistic continuous time, stochastic processes, including exponential Lévy models. We formally confirm many of the DCA properties that have been documented in the literature only by empirical studies, and provide mathematical insights into the risk, risk-return, and related performance metrics. A fundamental distributional equivalence linking DCA to arithmetic averaging of the underlying is obtained, which can be used to calculate risk.

We theoretically prove and demonstrate, with computational experiments and an empirical study of the S&P500 from 1954 to 2019, that the frequency of DCA investment M has a fundamental impact on risk, return and risk-return trade-offs. This fact that the fundamental characteristic M of DCA has been typically overlooked in most empirical (and theoretical) studies of DCA, as it was implicitly assumed to be insignificant, emphasises the importance and counterintuitiveness of this result. Consequently, DCA performance analyses need to take into account M, risk, return and risk-return trade-offs to provide balanced comparison to other investment strategies. We also provide a general computational framework for calculating risk and performance measures for different market timing strategies.

In terms of future work, we would like to investigate the impact of transaction costs upon the investment strategy and risk. This is particularly pertinent given that M plays a significant role in the DCA and related investment strategies. Other interesting questions[16] are to investigate how much we lose by doing DCA and most importantly the source of this loss in the characteristics of the traded assets (mean and/or volatility). We would also like to investigate the impact of stochastic interest rates, since interest rates fluctuate over time and DCA related strategies may differ over different time periods. For example, the inclusion of regime-switching

[16] We would like to thank one of the referees for suggesting these ideas.

interest rates (and market returns) is a parsimonious way to extend the analysis. Finally, we would like to investigate portfolio investment with respect to DCA.

References

Ahmadi-Javid, A. and M. Fallah-Tafti (2019). Portfolio optimization with entropic value-at-risk. *European Journal of Operational Research* 279(16), 225–241.

Artzner, P., F. Delbaen, J. Eber, D. Heath, and H. Ku (2003). Coherent multiperiod risk measurement. *Manuscript, ETH Zurich* .

Babat, O., J. Vera, and L. Zuluaga (2018). Computing near-optimal Value-at-Risk portfolios using integer programming techniques. *European Journal of Operational Research* 266(1), 304–315.

Bachelier, L. (1900). *Théorie de la spéculation.* Gauthier-Villars.

Balvers, R. J. and D. W. Mitchell (2000). Efficient gradualism in intertemporal portfolios. *Journal of Economic Dynamics and Control* 24(1), 21–38.

Barndorff-Nielsen, O. E. (1997). Processes of normal inverse gaussian type. *Finance and stochastics* 2(1), 41–68.

Battauz, A., M. De Donno, and A. Sbuelz (2015). Kim and omberg revisited: the duality approach. *Journal of Probability and Statistics* pp. 1–6.

Battauz, A., M. De Donno, and A. Sbuelz (2017). Reaching nirvana with a defaultable asset? *Decisions in Economics and Finance* 40, 31–52.

Battauz, A., M. De Donno, and A. Sbuelz (2018). On the exercise of american quanto options. *Working Paper* .

Battauz, A. and F. Rotondi (2019). American options and stochastic interest rates. *Working Paper* .

Battauz, A. and A. Sbuelz (2018). Non-myopic portfolio choice with unpredictable returns: the jump-to-default case. *European Financial Management* 24, 192–208.

Beck, A. (2014). *Introduction to nonlinear optimization: theory, algorithms, and applications with MATLAB*, volume 19. Siam.

Bi, J., H. Jin, and Q. Meng (2018). Behavioral mean-variance portfolio selection. *European Journal of Operational Research* 271(2), 644–663.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* 81(3), 637–654.

Boguth, O. and M. Simutin (2018). Leverage constraints and asset prices: Insights from mutual fund risk taking. *Journal of Financial Economics* 127(2), 325–341.

Boyarchenko, S. and S. Levendorskii (2000). Option pricing for truncated Levy processes. *International J. of Theoretical and Applied Finance* 3, 549–552.

Boyarchenko, S. and S. Levendorskii (2002). *Non-Gaussian Merton-Black-Scholes Theory.* Adv. Ser. Stat. Sci. Appl. Probab. World Scientific Publishing Co., River Edge, NJ,, volume 9 edition.

Boyd, S. and L. Vandenberghe (2004). *Convex optimization.* Cambridge university press.

Brandtner, M., W. Kursten, and R. Rischau (2018). Entropic risk measures and their comparative statics in portfolio selection: Coherence vs. convexity. *European Journal of Operational Research* 264(2), 707–716.

Braselton, J., J. Rafter, P. Humphrey, and M. Abell (1999). Randomly walking through wall street. *Mathematics and Computers in Simulation* 49(4-5), 297–318.

Brennan, M. J., F. Li, and W. N. Torous (2005). Dollar cost averaging. *Review of Finance* 9(4), 509–535.

Campbell, J. R. and Z. Hercowitz (2019). Liquidity constraints of the middle class. *American Economic Journal: Economic Policy* 11(3), 130–155.

Carr, P., H. Geman, D. B. Madan, and M. Yor (2002). The fine structure of asset returns: An empirical investigation. *The Journal of Business* 75(2), 305–332.

Constantinides, G. M. (1979). A note on the suboptimality of dollar-cost averaging as an investment policy. *Journal of Financial and Quantitative Analysis* 14(2), 443–450.

Cui, X., J. Gao, Y. Shi, and S. Zhu (2019a). Time-consistent and self-coordination strategies for multi-period mean-Conditional Value-at-Risk portfolio selection. *European Journal Operational Research* 276(2), 781–789.

Cui, Z., J. Kirkby, and D. Nguyen (2018). Efficient simulation of stochastic differential equations based on Markov chain approximations with applications. *Working Paper* .

Cui, Z., J. Kirkby, and D. Nguyen (2019b). A general framework for time-changed Markov processes and applications. *European Journal of Operational Research* 273(2), 785–800.

Cui, Z., J. Kirkby, and D. Nguyen (2019c). Nonparametric density estimation by B-spline duality. *Econometric Theory* pp. 1–42.

Cui, Z., J. L. Kirkby, and D. Nguyen (2017). A general framework for discretely sampled realized variance derivatives in stochastic volatility models with jumps. *European Journal of Operational Research* 262(1), 381–400.

Cuthbertson, K., D. Nitzsche, and N. Osullivan (2010). The market timing ability of uk mutual funds. *Journal of Business Finance & Accounting* 37(1-2), 270–289.

Dodson, L. (1989). Personal financial planning: Three dimensions of diversification. *Journal of Accountancy* 167(6), 131.

Dubil, R. (2005). Lifetime dollar-cost averaging: Forget cost savings, think risk reduction. *Journal of financial planning* 18(10).

Fong, W. M. (2017). Beating the market: Dollar-cost averaging with the profitable dividend yield strategy. *The Journal of Wealth Management* 20(2), 54–66.

Grable, J. E. and S. Chatterjee (2015). Another look at lump-sum versus dollar-cost averaging. *Journal of Financial Service Professionals* 69(5), 16–18.

Israelsen, C. L. (1999). Lump sums take their lumps. *Financial Planning* 29(1), 51–54.

Kirkby, J. (2015). Efficient option pricing by frame duality with the fast Fourier transform. *SIAM Journal on Financial Mathematics* 6(1), 713–747.

Kirkby, J. (2016). An efficient transform method for Asian option pricing. *SIAM Journal on Financial Mathematics* 7(1), 845–892.

Kirkby, J. (2017). Robust option pricing with characteristic functions and the B-spline order of density projection. *Journal of Computational Finance* 21(2), 101–127.

Kirkby, J. (2018). American and exotic option pricing with jump diffusions and other Levy processes. *Journal of Computational Finance* 22(3), 89–148.

Kirkby, J. and S. Deng (2019). Static hedging and pricing of exotic options with payoff frames. *Mathematical Finance* 29(2), 612–658.

Kirkby, J. and D. Nguyen (2020). Efficient Asian option pricing under regime switching jump diffusions and stochastic volatility models. *Annals of Finance* .

Knight, J. R. and L. Mandell (1992). Nobody gains from dollar cost averaging analytical, numerical and empirical results. *Financial Services Review* 2(1), 51–61.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management science* 48(8), 1086–1101.

Leggio, K. B. and D. Lien (2003). An empirical examination of the effectiveness of dollar-cost averaging using downside risk performance measures. *Journal of Economics and Finance* 27(2), 211.

Leitao, A., J. Kirkby, and L. Ortiz-Gracia (2019). The CTMC-Heston model: calibration and exotic option pricing with SWIFT. *Working Paper* .

Ling, A., J. Sun, and M. Wang (2019). Robust multi-period portfolio selection based on downside risk with asymmetrically distributed uncertainty set. *European Journal of Operational Research* .

Luskin, J. M. (2017). Dollar-cost averaging using the cape ratio: An identifiable trend influencing outperformance. *Journal of Financial Planning* 30(1), 54–60.

Lwin, K., R. Qu, and B. MacCarthy (2017). Mean-VaR portfolio optimization: A nonparametric approach. *European Journal of Operational Research* 260(2), 751–766.

Ma, G., C. Siu, and S.-P. Zhu (2019). Dynamic portfolio choice with return predictability and transaction costs. *European Journal of Operational Research* 3(1), 976–988.

Madan, D. B., P. P. Carr, and E. C. Chang (1998). The variance gamma process and option pricing. *Review of Finance* 2(1), 79–105.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance* 7(1), 77–91.

Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3(1-2), 125–144.

Milevsky, M. A. and S. E. Posner (2003). A continuous-time reexamination of dollar-cost averaging. *International Journal of Theoretical and Applied Finance* 6(02), 173–194.

Panyagometh, K. and K. X. Zhu (2016). Dollar-cost averaging, asset allocation, and lump sum investing. *The Journal of Wealth Management* 18(4), 75–89.

Penev, S., P. Shevchenko, and W. Wu (2019). The impact of model risk on dynamic portfolio selection under multi-period mean-standard-deviation criterion. *European Journal of Operational Research* 272(2), 772–784.

Pun, C. and H. Wong (2019). A linear programming model for selection of sparse high-dimensional multiperiod portfolios. *European Journal of Operational Research* 273(2), 754–771.

Pye, G. (1971). Minimax policies for selling an asset and dollar averaging. *Management Science* 17(7), 379–393.

Rubin, H. and C. Spaht (2018). Quality Dollar Cost Averaging Investing Versus Quality Index Investing. *Journal of Applied Business and Economics* 20(6).

Smith, G. and H. M. Artigue (2018). Another look at dollar cost averaging. *The Journal of Investing* 27(2), 66–75.

Staino, A. and E. Russo (2019). Nested Conditional Value-at-Risk portfolio selection: A model with temporal dependence driven by market-index volatility. *European Journal of Operational Research* 280(2), 741–753.

Szegö, G. (2005). Measures of risk. *European Journal of Operational Research* 163(1), 5–19.

Telyukova, I. A. (2013). Household need for liquidity and the credit card debt puzzle. *The Review of Economic Studies* 80(3), 1148–1177.

Thorley, S. (1994). The fallacy of dollar cost averaging. *Financial Practice & Education* 4(2), 138–143.

Tomlinson, L. E. (2012). *Successful Investing Formulas.* Barron's Publishing Company.

Trainor, W. J. (2005). Within-horizon exposure to loss for dollar cost averaging and lump sum investing. *Financial Services Reviews - Greenwich* 14(4), 319.

Vanduffel, S., A. Ahcan, L. Henrard, and M. Maj (2012). An explicit option-based strategy that outperforms dollar cost averaging. *International Journal of Theoretical and Applied Finance* 15(02), 1250013.

Wachter, J. (2010). Asset allocation. *Annual Review of Financial Economics* 2(1), 175–206.

Wang, W. and Z. Zhang (2019). Computing the gerber-shiu function by frame duality projection. *Scandanavian Actuarial Journal* 2019(4), 291–307.

Williams, R. E. and P. W. Bacon (1993). Lump sum beats dollar-cost averaging. *Journal of Financial Planning* 6(2).

Zhang, Z., Y. Yong, and W. Yu (2020). Valuing equity-linked death benefits in general exponential Lévy models. *Journal of Computational and Applied Mathematics* 365, 112377.

A Characteristic exponent

Model	$\psi_L(\xi)$	$\psi_R(-2i)$
BSM	$-\frac{\sigma^2}{2}\xi^2$	$\zeta+\sigma^2$
MJD	$-\frac{\sigma^2}{2}\xi^2+\lambda\Big(\exp(\mathrm{i}\xi\mu_J-\frac{\sigma_J^2}{2}\xi^2)-1\Big)$	$\zeta+\sigma^2+\lambda[\exp\left(2(\mu_J+\sigma_J^2)\right)-2\exp(\mu_J+\frac{1}{2}\sigma_J^2)+1]$
CGMY	$C\Gamma(-Y)\left((M-\mathrm{i}\xi)^Y-M^Y+(G+\mathrm{i}\xi)^Y-G^Y\right)$	$\zeta+C\Gamma(-Y)\left((M-2)^Y-2(M-1)^Y+M^Y+(G+2)^Y-2(G+1)^Y+G^Y\right)$
NIG	$-\delta\left(\sqrt{\alpha^2-(\beta+\mathrm{i}\xi)^2}-\sqrt{\alpha^2-\beta^2}\right)$	$\zeta-\delta\left(\sqrt{\alpha^2-(\beta-2)^2}-\sqrt{\alpha^2-\beta^2}\right)$
Kou	$-\frac{\sigma^2}{2}\xi^2+\lambda\Big(\frac{p\eta_1}{\eta_1-\mathrm{i}\xi}+\frac{(1-p)\eta_2}{\eta_2+i\xi}-1\Big)$	$\zeta+2\sigma^2+\lambda\Big(\frac{p\eta_1}{\eta_1+2}+\frac{(1-p)\eta_2}{\eta_2-2}-1\Big)$
VG	$-\frac{\sigma^2}{2}\xi^2-\frac{1}{\nu}\log\left(1-i\nu\theta\xi+\nu\frac{\sigma_V^2}{2}\xi^2\right)$	$\zeta+2\sigma^2+\frac{1}{\nu}\log\left(1+2(\nu\theta-\nu\sigma_V^2)\right)$

Table 3: Characteristic exponents (Lévy symbols) $\psi_L(\xi)$ and $\psi_R(-2i)$ for some models. Here we denote the convexity-corrected drift by $\zeta := 2(\mu-\psi_L(-i))$, where $\psi_L(-i)$ can be computed in closed-form from the first column.

B Proofs

Proof of Lemma 2: Noting that $\mathbb{E}\left[S_T/S_{t_m}\right]=\mathbb{E}\left[\mathbb{E}[S_T/S_{t_m}]|S_{t_m}\right]]=\mathbb{E}[e^{\mu(T-t_m)}]=e^{\mu(T-t_m)}$. Using Lemma 1, we have

$$\begin{aligned}\mathbb{E}\left[W_T(\boldsymbol{\alpha})\right]&=\sum_{m=0}^{M}\alpha_m\mathbb{E}\left[\frac{S_T}{S_{t_m}}\right]+e^{rT}W-\sum_{k=0}^{M}\alpha_k e^{r(T-t_k)}\\&=\sum_{m=0}^{M}\alpha_m e^{\mu(T-t_m)}+e^{rT}W-\sum_{k=0}^{M}\alpha_k e^{r(T-t_k)},\end{aligned}$$

and (8) follows. The bounds in (9) follow from (8), together with the observation that $\left(e^{\mu(T-t_m)}-e^{r(T-t_m)}\right)$ is monotonically decreasing in m whenever $\mu\geq r$. The final statement follows by noting that if we shift any amount of wealth from the final period, α_M, to any α_m where $m<M$, $E\left[W_T(\boldsymbol{\alpha})\right]$ will increase, again by monotonicity, and the worst expected outcome of $e^{rT}W$ is attained when $\alpha_M=W$. □

Proof of Lemma 3: First note that

$$\mathbb{E}\left[e^{\theta R_t}\right]=e^{\psi_R(-i\theta)t}=e^{i\mu(-i\theta)t+\psi_L(-i\theta)t}=e^{\mu\theta t+\psi_L(-i\theta)t},$$

and when $\theta = 1$, $\mathbb{E}\left[e^{R_t}\right] = e^{\psi_R(-i\theta)t} = e^{\mu t}$. Let $R_m := \log(S_{t_m}/S_{t_{m-1}})$, for $m = 1, \ldots, M$. Let $X_m := \exp(R_{m+1} + \cdots + R_M)$, we have

$$\begin{aligned}
Var(W_T(\boldsymbol{\alpha})) &= Var\left(\sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}}\right) = Var\left(\sum_{m=0}^{M-1} \alpha_m \frac{S_T}{S_{t_m}}\right) && (46)\\
&= Var\left(\sum_{m=0}^{M-1} \alpha_m \exp(R_{m+1} + \cdots + R_M)\right)\\
&= Var\left(\sum_{m=0}^{M-1} \alpha_m X_m\right)\\
&= \sum_{m=0}^{M-1} \alpha_m^2 Var(X_m) + 2 \sum_{0 \le m < j \le M-1} \alpha_m \alpha_j Cov(X_m, X_j)\\
&= \sum_{m=0}^{M-1} \alpha_m^2 Var(X_m) + 2 \sum_{0 \le m < j \le M-1} \alpha_m \alpha_j (\mathbb{E}[X_m X_j] - \mathbb{E}[X_m]\mathbb{E}[X_j]). && (47)
\end{aligned}$$

For $k \geq 1$, with $m + k \leq M$, by independence of increments

$$\begin{aligned}
\mathbb{E}[X_m X_{m+k}] &= \mathbb{E}\left[\exp\left(\sum_{j=m+1}^{m+k} R_j + \sum_{j=m+k+1}^{M} R_j\right)\right]\\
&= \mathbb{E}\left[\exp\left(\sum_{j=m+1}^{m+k} R_j\right)\right] \mathbb{E}\left[\exp\left(2 \sum_{j=m+k+1}^{M} R_j\right)\right]\\
&= \exp(k\Delta t \cdot \psi_R(-i)) \exp((M - m - k)\Delta t \cdot \psi_R(-2i)),
\end{aligned}$$

and note that we have

$$\mathbb{E}[X_m] = \mathbb{E}[\exp(\sum_{i=m+1}^{M} R_i)] = \exp((M - m)\Delta t \cdot \psi_R(-i)).$$

Moreover,

$$\begin{aligned}
Var(X_m) &= \mathbb{E}[(\prod_{i=m+1}^{M} \exp(R_i))^2] - (\mathbb{E}(\prod_{i=m+1}^{M} \exp(R_i)))^2\\
&= \mathbb{E} \prod_{i=m+1}^{M} \exp(2R_i) - \prod_{i=m+1}^{M} \left(\mathbb{E}(\exp(R_i))\right)^2\\
&= \mathbb{E} \exp(\sum_{i=m+1}^{M} 2R_i) - \prod_{i=m+1}^{M} \left(\mathbb{E}(\exp(R_i))\right)^2\\
&= \exp((M - m)\Delta t \cdot \psi_R(-2i)) - \prod_{i=m+1}^{M} \exp(2\Delta t \cdot \psi_R(-i))\\
&= \exp((M - m)\Delta t \cdot \psi_R(-2i)) - \exp(2(M - m)\Delta t \cdot \psi_R(-i)).
\end{aligned}$$

Substituting all in (46), we have

$$
\begin{aligned}
&Var\left(\sum_{m=0}^{M-1} \alpha_m \frac{S_T}{S_{t_m}}\right)\\
&= \sum_{m=0}^{M-1} \alpha_m^2 \left[\exp((M-m)\Delta t \cdot \psi_R(-2i)) - \exp(2(M-m)\Delta t \cdot \psi_R(-i))\right] +\\
&+ 2 \sum_{0 \le m,k \le M-1, m+k \le M} \alpha_m \alpha_{m+k} \left[\exp(k\Delta t \cdot \psi_R(-i)) \exp((M-m-k)\Delta t \cdot \psi_R(-2i))\right]\\
&- 2 \sum_{0 \le m,k \le M-1, m+k \le M} \alpha_m \alpha_{m+k} \left[\exp((2M-2m-k)\Delta t \cdot \psi_R(-i))\right]. \qquad (48)
\end{aligned}
$$

Noting that $\psi_R(-i) = \mu$, the result follows. □

Proof of Theorem 1: Recall that the number of shares purchased at time t_m is given by

$$n_m = \frac{\alpha_m}{S_{t_m}} = \frac{W}{(M+1) \cdot S_{t_m}}, \quad m = 0, 1, \ldots, M.$$

Hence we have, the average cost is given by

$$
\begin{aligned}
\frac{W}{n_0 + n_1 + \ldots + n_M} &= \frac{W}{\frac{W}{M+1} \sum_{m=0}^{M} \frac{1}{S_{t_m}}}\\
&= \frac{M+1}{\sum_{m=0}^{M} \frac{1}{S_{t_m}}} = H(S_{t_0}, S_{t_1}, \ldots, S_{t_M}), \qquad (49)
\end{aligned}
$$

where $H(x_0, x_2, \ldots, x_M)$ denotes the Harmonic mean of $x_0, x_2, \ldots, x_M$. It is well known that the Harmonic mean is less than or equal to arithmetic mean. That is,

$$H(x_0, x_2, \ldots, x_M) \le \frac{1}{M+1} \sum_{m=0}^{M} x_m.$$

This completes the proof of the theorem. □

Proof of Corollary 2: Applying (8) with $t_m = m\Delta t = Tm/M$, and $\alpha_m = W/(M+1)$

$$
\begin{aligned}
\mathbb{E}\left[W_T^{DCA}\right] &= e^{rT} W + \frac{W}{M+1} \sum_{m=0}^{M-1} \left(e^{\mu T(1-m/M)} - e^{rT(1-m/M)}\right)\\
&\overset{m \to M-m}{=} e^{rT} W + \frac{W}{M+1} \sum_{m=1}^{M} \left(e^{\mu T(m/M)} - e^{rT(m/M)}\right)\\
&= e^{rT} W + \frac{W}{M+1} \sum_{m=0}^{M} \left(e^{\mu T(m/M)} - e^{rT(m/M)}\right)\\
&= e^{rT} W + \frac{W}{M+1} \left(\frac{e^{\mu T(M+1)/M} - 1}{e^{\mu T/M} - 1} - \frac{e^{rT(M+1)/M} - 1}{e^{rT/M} - 1}\right).
\end{aligned}
$$

Next, for the limit we have

$$
\begin{aligned}
\mathbb{E}\left[W_T^{DCA}\right] &= e^{rT} W + \frac{W}{M+1} \sum_{m=0}^{M} \left(e^{\mu T(m/M)} - e^{rT(m/M)}\right)\\
&\overset{M \to \infty}{\longrightarrow} e^{rT} W + W \int_0^1 (e^{\mu T x} - e^{rTx}) dx = e^{rT} W + W \left(\frac{e^{\mu T} - 1}{\mu T} - \frac{e^{rT} - 1}{rT}\right).
\end{aligned}
$$

For the monotonicity, to simplify notation, let $a = \mu T$ and $b = rT$, and $W = 1$, we have

$$\begin{aligned}
\mathbb{E}\left[W_T^{DCA}\right] &= e^{rT} + \frac{1}{M+1}\sum_{m=0}^{M}\left(e^{a(m/M)} - e^{b(m/M)}\right) \\
&= e^{rT} + \frac{1}{M+1}\sum_{m=0}^{M-1}\left(e^{a(m/M)} - e^{b(m/M)}\right) + \frac{1}{M+1}(e^a - a^b) \\
&= e^{rT} + \frac{1}{M+1}\left(\frac{e^a - 1}{e^{a/M} - 1} - \frac{e^b - 1}{e^{b/M} - 1}\right) + \frac{1}{M+1}(e^a - a^b). \qquad (50)
\end{aligned}$$

From (50), taking the derivative with respect to M, we have

$$\begin{aligned}
\frac{d}{dM}\mathbb{E}\left[W_T^{DCA}\right] &= \frac{1}{M^2(M+1)}\left[\frac{a(e^a-1)e^{a/M}}{(e^{a/M}-1)^2} - \frac{b(e^b-1)e^{b/M}}{(e^{b/M}-1)^2}\right] \\
&\quad - \frac{1}{(M+1)^2}\left[\frac{(e^a-1)}{(e^{a/M}-1)} - \frac{(e^b-1)}{(e^{b/M}-1)}\right] - \frac{e^a - e^b}{(M+1)^2} \\
&= \frac{1}{M+1}\Bigg(\frac{1}{M^2}\left[\frac{a(e^a-1)e^{a/M}}{(e^{a/M}-1)^2} - \frac{b(e^b-1)e^{b/M}}{(e^{b/M}-1)^2}\right] \\
&\quad - \frac{1}{(M+1)}\left[\frac{(e^a-1)}{(e^{a/M}-1)} - \frac{(e^b-1)}{(e^{b/M}-1)}\right] - \frac{e^a - e^b}{(M+1)}\Bigg). \qquad (51)
\end{aligned}$$

Now let $f(x)$ be the following function

$$f(x) = \frac{1}{M^2}\cdot\frac{x(e^x-1)e^{x/M}}{(e^{x/M}-1)^2} - \frac{1}{M+1}\cdot\frac{e^x - 1}{e^{x/M}-1} - \frac{e^x}{M+1}, \quad x > 0.$$

Since the term $\frac{-e^x}{M+1}$ is dominating, it can be checked that $f(x)$ is decreasing[17] for all $x > 0$. In particular, since $a \geq b$, $f(a) \leq f(b)$. As a result, from (51), it can be seen that $\frac{d}{dM}\mathbb{E}\left[W_T^{DCA}\right] \leq 0$ for all M. As a result, $\mathbb{E}\left[W_T^{DCA}\right]$ is decreasing in term of M. This completes the proof of the corollary. □

Proof of Corollary 3: Without loss of generality, we can assume that $W = 1$. From Lemma 3 we have

$$\begin{aligned}
(M+1)^2 Var\left[W_T^{DCA}\right] &= \sum_{m=0}^{M-1}\left[\exp((M-m)\Delta t\cdot\psi_R(-2i)) - \exp(2(M-m)\Delta t\cdot\mu)\right] + \\
&\quad + 2\sum_{m=0}^{M-2}\sum_{k=1}^{M-m-1}\left[\exp(k\Delta t\cdot\mu)\exp((M-m-k)\Delta t\cdot\psi_R(-2i))\right] \\
&\quad - 2\sum_{m=0}^{M-2}\sum_{k=1}^{M-m-1}\left[\exp((2(M-m)-k)\Delta t\cdot\mu)\right]. \qquad (52)
\end{aligned}$$

[17]This is obvious for large M, for small M it can be seen from plotting the derivative of $f(x)$.

Consider the first term of (52),

$$\begin{aligned}
\Pi_1 &:= \sum_{m=0}^{M-1} \left[\exp((M-m)\Delta t \cdot \psi_R(-2i)) - \exp(2(M-m)\Delta t \cdot \mu)\right] \\
&\overset{m \to M-m}{=} \sum_{m=1}^{M} \left[\exp(m\Delta t \cdot \psi_R(-2i)) - \exp(2m\Delta t \cdot \mu)\right] \\
&= \sum_{m=0}^{M} \left[\exp(m\Delta t \cdot \psi_R(-2i)) - \exp(2m\Delta t \cdot \mu)\right] \\
&= \frac{\exp((M+1)\Delta t \cdot \psi_R(-2i)) - 1}{\exp(\Delta t \cdot \psi_R(-2i)) - 1} - \frac{\exp(2(M+1)\Delta t \cdot \mu) - 1}{\exp(2\Delta t \cdot \mu) - 1}.
\end{aligned} \tag{53}$$

Next, consider the second term of (52), we have

$$\begin{aligned}
\Pi_2 &:= \sum_{m=0}^{M-2} \sum_{k=1}^{M-m-1} \left[\exp(k\Delta t \cdot \mu) \exp((M-m-k)\Delta t \cdot \psi_R(-2i))\right] \\
&= \sum_{m=0}^{M-2} \sum_{k=1}^{M-m-1} \exp((M-m)\Delta t \cdot \psi_R(-2i)) \exp(k\Delta t \cdot (\mu - \psi_R(-2i))) \\
&= \sum_{m=0}^{M-2} \Bigg(\exp((M-m)\Delta t \cdot \psi_R(-2i)) \exp(\Delta t \cdot (\mu - \psi_R(-2i))) \\
&\qquad \cdot \frac{\exp((M-m-1)\Delta t \cdot (\mu - \psi_R(-2i))) - 1}{\exp(\Delta t \cdot (\mu - \psi_R(-2i))) - 1} \Bigg) \\
&= \frac{\exp(\Delta t \cdot (\mu - \psi_R(-2i)))}{\exp(\Delta t(\mu - \psi_R(-2i))) - 1} \\
&\quad \cdot \sum_{m=0}^{M-2} \exp((M-m)\Delta t \psi_R(-2i)) \left(\exp((M-m-1)\Delta t \cdot (\mu - \psi_R(-2i))) - 1\right).
\end{aligned} \tag{54}$$

In the equation (54) above, the term

$$\begin{aligned}
&\sum_{m=0}^{M-2} \exp((M-m)\Delta t \cdot \psi_R(-2i)) \left(\exp((M-m-1)\Delta t \cdot (\mu - \psi_R(-2i))) - 1\right) \\
&= \exp(\Delta t \psi_R(-2i)) \sum_{m=0}^{M-2} \exp((M-m-1)\Delta t \cdot \mu) - \sum_{m=0}^{M-2} \exp((M-m)\Delta t \cdot \psi_R(-2i)) \\
&= \exp(\Delta t \psi_R(-2i)) \sum_{m=1}^{M-1} \exp(m\Delta t \cdot \mu) - \exp(2 \cdot \Delta t \cdot \psi_R(-2i)) \sum_{m=0}^{M-2} \exp(m\Delta t \cdot \psi_R(-2i)) \\
&= \exp(\Delta t(\mu + \psi_R(-2i))) \frac{\exp((M-1)\Delta t \cdot \mu) - 1}{\exp(\Delta t \cdot \mu) - 1} \\
&\qquad - \exp(2 \cdot \Delta t \cdot \psi_R(-2i)) \frac{\exp((M-1)\Delta t \cdot \psi_R(-2i)) - 1}{\exp(\Delta t \cdot \psi_R(-2i)) - 1}.
\end{aligned}$$

Lastly, consider the third term of (52), similarly to the above we have

$$
\begin{aligned}
\Pi_3 &:= \sum_{m=0}^{M-2} \sum_{k=1}^{M-m-1} [\exp((2(M-m)-k)\Delta t \cdot \mu)] \\
&= \sum_{m=0}^{M-2} \exp(2(M-m)\Delta t \cdot \mu) \sum_{k=1}^{M-m-1} \exp(-2k\Delta t \cdot \mu) \\
&= \sum_{m=0}^{M-2} \exp(2(M-m)\Delta t \cdot \mu) \exp(-\Delta t \cdot \mu) \frac{\exp(-(M-m-1)\Delta t \cdot \mu) - 1}{\exp(-\Delta t \cdot \mu) - 1} \\
&= \frac{\exp(-\Delta t \cdot \mu)}{\exp(-\Delta t \cdot \mu) - 1} \sum_{m=0}^{M-2} (\exp((M-m+1)\Delta t\mu) - \exp(2(M-m)\Delta t \cdot \mu)) \\
&= \frac{\exp(-\Delta t \cdot \mu)}{\exp(-\Delta t \cdot \mu) - 1} \left(\exp(3\Delta t\mu) \frac{\exp((M-1)\Delta t\mu) - 1}{\exp(\Delta t\mu) - 1} - \exp(4\Delta t \cdot \mu) \frac{\exp(2(M-1)\Delta t \cdot \mu)) - 1}{\exp(2\Delta t \cdot \mu) - 1} \right).
\end{aligned} \tag{55}
$$

By combining (53), (54), and (55), we can write the variance as a function in term of M. That is

$$
Var\left[W_T^{DCA}\right] = \frac{1}{(M+1)^2} (\Pi_1 + 2\Pi_2 - 2\Pi_3) \geq 0. \tag{56}
$$

□

Proof of Corollary 4: Without loss of generality, we can assume that $W = 1$. Recall from Corollary 3 that

$$
Var\left[W_T^{DCA}\right] = \frac{W^2}{(M+1)^2} (\Pi_1 + 2\Pi_2 - 2\Pi_3),
$$

where definitions of Π_1, Π_2, and Π_3 are there defined. Recall that $\Delta t = T/M$, we then have

$$
\begin{aligned}
\frac{1}{(M+1)^2}\Pi_1 &= \frac{1}{(M+1)^2} \left(\frac{\exp((M+1)/M \cdot T\psi_R(-2i)) - 1}{\exp(T/M \cdot \psi_R(-2i)) - 1} - \frac{\exp(2(M+1)/M \cdot T\mu) - 1}{\exp(2T/M \cdot \mu) - 1} \right) \\
&\sim \frac{1}{(M+1)^2} \left(\frac{\exp((M+1)/M \cdot T\psi_R(-2i)) - 1}{T/M \cdot \psi_R(-2i)} - \frac{\exp(2(M+1)/M \cdot T\mu) - 1}{2T/M \cdot \mu} \right) \\
&= \frac{M}{(M+1)^2} \left(\frac{\exp((M+1)/M \cdot T\psi_R(-2i)) - 1}{T\psi_R(-2i)} - \frac{\exp(2(M+1)/M \cdot T\mu) - 1}{2T\mu} \right) \\
&\longrightarrow 0 \quad \text{as} \quad M \to \infty.
\end{aligned} \tag{57}
$$

Similarly,

$$
\begin{aligned}
\frac{\Pi_2}{(M+1)^2} &\sim \frac{M^2}{(M+1)^2} \frac{\exp(T/M(\mu - \psi_R(-2i)))}{T(\mu - \psi_R(-2i))} \Big(\exp(T/M(\mu + \psi_R(-2i))) \frac{\exp((M-1)/M \cdot T\mu) - 1}{T\mu} \\
&\quad - \exp(2T/M\psi_R(-2i)) \frac{\exp((M-1)/M \cdot T\psi_R(-2i)) - 1}{T\psi_R(-2i)} \Big) \\
&\longrightarrow \frac{1}{T(\mu - \psi_R(-2i))} \left(\frac{\exp(T\mu) - 1}{T\mu} - \frac{\exp(T\psi_R(-2i)) - 1}{T\psi_R(-2i)} \right).
\end{aligned} \tag{58}
$$

Lastly, we have

$$
\begin{aligned}
\frac{\Pi_3}{(M+1)^2} &\sim \frac{M^2}{(M+1)^2}\frac{\exp(-T/M\mu)}{-T\mu}\Big(\exp(3T/M\mu)\frac{\exp((M-1)/M\cdot T\mu)-1}{T\mu} \\
&\quad - \exp(4T/M\mu)\frac{\exp(2(M-1)/M\cdot T\mu)-1}{2T\mu}\Big) \\
&\longrightarrow \frac{-1}{T\mu}\left(\frac{\exp(T\mu)-1}{T\mu} - \frac{\exp(2T\mu)-1}{2T\mu}\right) = \frac{-1}{2(T\mu)^2}\left(2\exp(T\mu)-\exp(2T\mu)-1\right).
\end{aligned} \tag{59}
$$

As a result, we have as $M\to\infty$:

$$
\begin{aligned}
Var\left[W_T^{DCA}\right] \longrightarrow\ & \frac{2}{T(\mu-\psi_R(-2i))}\left(\frac{\exp(T\mu)-1}{T\mu} - \frac{\exp(T\psi_R(-2i))-1}{T\psi_R(-2i)}\right) \\
& + \frac{1}{(T\mu)^2}\left(2\exp(T\mu)-\exp(2T\mu)-1\right).
\end{aligned}
$$

From (57),(58) and (59), it can be seen that $\frac{1}{(M+1)^2}\Pi_i$, for $i=1,2,3$ is bounded. Therefore, the variance $Var\left[W_T^{DCA}\right]$ is bounded and is dominated by $\frac{1}{(M+1)^2}\Pi_1$.

Let

$$
\begin{aligned}
f(M) &= \big(\exp(\psi_R(-2i)T)-\exp(2\mu T)\big) - Var\left[W_T^{DCA}\right] \\
&= \big(\exp(\psi_R(-2i)T)-\exp(2\mu T)\big) - \frac{1}{(M+1)^2}\left(\Pi_1+2\Pi_2-2\Pi_3\right).
\end{aligned}
$$

Since $Var\left[W_T^{DCA}\right]$ is positive, bounded and is dominated by $\frac{1}{(M+1)^2}\Pi_1$, it can be checked that $f(M)$ is a positive decreasing function. Hence the proof of the theorem follows immediately. $\square$

Proof of Theorem 6: Since the matrix H is symmetric and positive definite, the solution to (23) is guranteed to exist, see Boyd and Vandenberghe (2004). By Theorem 11.5 in Beck (2014), we can next form the Lagrangian

$$
\begin{aligned}
L &= \frac{1}{2}Var\left(W_T(\boldsymbol{\alpha})\right) - \gamma_1\left(\mathbb{E}[W_T(\boldsymbol{\alpha})]-\widehat{\mu}\right) - \gamma_2\left(\sum_{m=0}^{M-1}\alpha_m - W\right) \\
&= \frac{1}{2}\boldsymbol{\alpha}^T H\boldsymbol{\alpha} - \gamma_1\left(e^{rT}W + \sum_{m=0}^{M-1}\alpha_m\left(e^{\mu(T-t_m)}-e^{r(T-t_m)}\right)-\widehat{\mu}\right) - \gamma_2\left(\sum_{m=0}^{M-1}\alpha_m - W\right)
\end{aligned}
$$

where γ_1,γ_2 are Lagrangian multipliers. We have, for $m=0,1\ldots,M-1$,

$$
\frac{\partial L}{\partial \alpha_m} = \sum_{i=1}^{M-1}\alpha_j H_{mj} - \gamma_1\left(e^{\mu(T-t_m)}-e^{r(T-t_m)}\right) - \gamma_2 = 0. \tag{60}
$$

$$
\frac{\partial L}{\partial \gamma_1} = e^{rT}W + \sum_{m=0}^{M-1}\alpha_m\left(e^{\mu(T-t_m)}-e^{r(T-t_m)}\right) - \widehat{\mu} = 0. \tag{61}
$$

$$
\frac{\partial L}{\partial \gamma_2} = \left(\sum_{m=0}^{M-1}\alpha_m - W\right) = 0. \tag{62}
$$

From (60), we have

$$\boldsymbol{\alpha}^* = H^{-1}\left(\gamma_1 \boldsymbol{y} + \gamma_2 \mathbf{1}\right). \tag{63}$$

where $\mathbf{1} = (1, 1, \ldots, 1)^T \in \mathbb{R}^M$ and $\boldsymbol{y} = \left(e^{\mu T} - e^{rT}, e^{\mu(T-t_1)} - e^{r(T-t_1)}, \ldots, e^{\mu(T-t_{M-1})} - e^{r(T-t_{M-1})}\right)^T$. The equation (61) gives

$$\begin{aligned}\widehat{\mu} - e^{rT}W &= \sum_{m=0}^{M-1} \alpha_m^* \left(e^{\mu(T-t_m)} - e^{r(T-t_m)}\right) = \boldsymbol{y}^T \boldsymbol{\alpha}^* \\ &= \boldsymbol{y}^T H^{-1}\left(\gamma_1 \boldsymbol{y} + \gamma_2 \mathbf{1}\right) \\ &= \gamma_1 \boldsymbol{y}^T H^{-1} \boldsymbol{y} + \gamma_2 \boldsymbol{y}^T H^{-1} \mathbf{1}.\end{aligned} \tag{64}$$

From (62), we have

$$W = \sum_{m=0}^{M-1} \alpha_m^* = \mathbf{1}^T \boldsymbol{\alpha}^* = \mathbf{1}^T H^{-1}\left(\gamma_1 \boldsymbol{y} + \gamma_2 \mathbf{1}\right) = \gamma_1 \mathbf{1}^T H^{-1} \boldsymbol{y} + \gamma_2 \mathbf{1}^T H^{-1} \mathbf{1}. \tag{65}$$

From (64) and (65), we have

$$\begin{cases} \gamma_1 = \dfrac{(\widehat{\mu} - e^{rT}W)\mathbf{1}^T H^{-1}\mathbf{1} - W\boldsymbol{y}^T H^{-1}\mathbf{1}}{(\boldsymbol{y}^T H^{-1}\boldsymbol{y})(\mathbf{1}^T H^{-1}\mathbf{1}) - (\boldsymbol{y}^T H^{-1}\mathbf{1})(\mathbf{1}^T H^{-1}\boldsymbol{y})}, \\ \gamma_2 = \dfrac{W\boldsymbol{y}^T H^{-1}\boldsymbol{y} - (\widehat{\mu} - e^{rT}W)\mathbf{1}^T H^{-1}\boldsymbol{y}}{(\boldsymbol{y}^T H^{-1}\boldsymbol{y})(\mathbf{1}^T H^{-1}\mathbf{1}) - (\boldsymbol{y}^T H^{-1}\mathbf{1})(\mathbf{1}^T H^{-1}\boldsymbol{y})}. \end{cases} \tag{66}$$

As a result, we have

$$\begin{aligned}\boldsymbol{\alpha}^* = H^{-1} &\left(\frac{(\widehat{\mu} - e^{rT}W)\mathbf{1}^T H^{-1}\mathbf{1} - W\boldsymbol{y}^T H^{-1}\mathbf{1}}{(\boldsymbol{y}^T H^{-1}\boldsymbol{y})(\mathbf{1}^T H^{-1}\mathbf{1}) - (\boldsymbol{y}^T H^{-1}\mathbf{1})(\mathbf{1}^T H^{-1}\boldsymbol{y})}\right)\boldsymbol{y} \\ + H^{-1} &\left(\frac{W\boldsymbol{y}^T H^{-1}\boldsymbol{y} - (\widehat{\mu} - e^{rT}W)\mathbf{1}^T H^{-1}\boldsymbol{y}}{(\boldsymbol{y}^T H^{-1}\boldsymbol{y})(\mathbf{1}^T H^{-1}\mathbf{1}) - (\boldsymbol{y}^T H^{-1}\mathbf{1})(\mathbf{1}^T H^{-1}\boldsymbol{y})}\right)\mathbf{1}.\end{aligned} \tag{67}$$

This completes the proof of the Theorem. □

Proof of Theorem 7: Let $R_m := \log(S_{t_m}/S_{t_{m-1}})$, for $m = 1, \ldots, M$. For an exponential Lévy process log returns are independent, and we have

$$\begin{aligned}\sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}} &= \sum_{m=0}^{M-1} \alpha_m \exp(R_{m+1} + \cdots + R_M) + \alpha_M \frac{S_T}{S_{t_M}} \\ &= \sum_{m=0}^{M-1} \alpha_m \exp\left(\sum_{j=m+1}^{M} R_j\right) + \alpha_M,\end{aligned}$$

where we note $R_j \overset{d}{=} R_k$ for $j, k = 1, \ldots, M$, due to uniform monitoring and the fact that R_j

are Lévy increments, and $(R_1,\ldots,R_M) \overset{d}{=} (R_M,\ldots,R_1)$. It then follows that

$$\begin{aligned}
\sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}} &= \sum_{m=0}^{M-1} \alpha_m \exp\left(\sum_{j=m+1}^{M} R_j\right) + \alpha_M \frac{S_T}{S_{t_M}} \\
&= \alpha_M + e^{R_M}\left[\alpha_{M-1} + e^{R_{M-1}}\left[\alpha_{M-2} + e^{R_{M-2}}\left[\cdots \alpha_2 + e^{R_2}\left[\alpha_1 + \alpha_0 e^{R_1}\right]\right]\right]\right] \\
&\overset{d}{=} \alpha_M + e^{R_1}\left[\alpha_{M-1} + e^{R_2}\left[\alpha_{M-2} + e^{R_3}\left[\cdots \alpha_2 + e^{R_{M-1}}\left[\alpha_1 + \alpha_0 e^{R_M}\right]\right]\right]\right] \\
&= \alpha_M \frac{S_0}{S_0} + \sum_{m=1}^{M} \alpha_{M-m} \exp\left(\sum_{j=1}^{m} R_j\right) \\
&= \frac{1}{S_0} \sum_{m=0}^{M} \alpha_{M-m} S_0 \exp\left(\sum_{j=1}^{m} R_j\right) \\
&= \frac{1}{S_0} \sum_{m=0}^{M} \alpha_{M-m} S_{t_m}.
\end{aligned}$$

In particular, we have proved the equivalence in distribution of the quantities

$$\sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}} \overset{d}{=} \frac{1}{S_0} \sum_{m=0}^{M} \alpha_{M-m} S_{t_m}. \tag{68}$$

The result then follows. $\square$

Proof of Corollary 8: Note that from (26)

$$\begin{aligned}
&\sum_{m=0}^{M} \alpha_m \frac{S_T}{S_{t_m}} \overset{d}{=} \frac{1}{S_0} \sum_{m=0}^{M} \alpha_{M-m} S_{t_m} \\
&= \left(\alpha_M + e^{R_1}\left(\alpha_{M-1} + e^{R_2}\left(\ldots e^{R_{M-1}}\left(\alpha_1 + \alpha_0 e^{R_M}\right)\right)\right)\right) \\
&= \alpha_M \left(1 + \frac{\alpha_{M-1}}{\alpha_M} e^{R_1}\left(1 + \frac{\alpha_{M-2}}{\alpha_{M-1}} e^{R_2}\left(\ldots \frac{\alpha_1}{\alpha_2} e^{R_{M-1}}\left(1 + \frac{\alpha_0}{\alpha_1} e^{R_M}\right)\right)\right)\right) \\
&= \alpha_M \left(1 + \exp\left(\log\left(\frac{\alpha_{M-1}}{\alpha_M}\right) + R_1 + \ldots \right.\right. \\
&\qquad\qquad \left.\left. + \log\left(R_{M-1} + \log\left(1 + \exp\left(\log\left(\frac{\alpha_0}{\alpha_1}\right) + R_M\right)\right)\right)\right)\right)
\end{aligned} \tag{69}$$

From the last equality in (69), let

$$\begin{aligned}
Y_1 &= \log(\alpha_0/\alpha_1) + R_M, \\
Y_m &= \log\left(\frac{\alpha_{m-1}}{\alpha_m}\right) + R_{M+1-m} + Z_{m-1}, \quad m = 2,\ldots,M.
\end{aligned}$$

we have

$$Y_m = \log\left(\frac{1}{\alpha_m S_{t_{M-m}}} \sum_{j=1}^{m} \alpha_{m-j} S_{t_{M-m+j}}\right), \quad m = 2,\ldots,M. \tag{70}$$

from which the equation (38) follows from Theorem 7. $\square$