Bootstrap resampling analysis of monthly stock log return

by

Yaxuan Guo

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computational Mathematics

Waterloo, Ontario, Canada, 2020

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In the financial area, researchers are interested in the asset allocation for a long time period. However, this allocation is based on limited historical observations. Thus, researchers try to enlarge the data set where bootstrap resampling, especially block bootstrap resampling, is the main method for samples with high correlation.

In this research, we derive the continuous form of two asset price models and test the performance of the stationary block bootstrap resampling in estimating the total log return in a given time horizon. Since we use the monthly log return along a sample path, we additionally analyse the bootstrap resampled results of the monthly return for reference. We find that the block bootstrap resampling performs well in estimating the statistics of the monthly return for both two models, however, the estimation for the total log return is less satisfactory. Since the log return of the Customized AR(1) model has significant auto-correlation, we also test the bootstrap resampled estimation of variance ratio in the continuous form of the Customized AR(1) model, compare it with that in the discrete form in [6] and find a same result that the bootstrap resampled bias is significant.

In [20], since we use the monthly return along a sample path, the distribution statistics such as mean, variance and the total log return is defined as the level-1 parameter, and the distribution statistics for level-1 parameter is called the level-2 parameter. Based on our computation, it is found that the block bootstrap resampling performs well in approximating the level-1 parameters, and it requires an optimal block size for distribution estimation of the level-1 parameters, which are the level-2 parameters.

Acknowledgements

I would like to thank Prof. Li and Prof. Forsyth for their guidance and support for this research, and Prof. George Labahn for reviewing this paper.

Dedication

This is dedicated to my family and friends for supporting me during my master degree.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

An interesting topic in the financial area is asset allocation for a long time horizon, e.g. 30 consecutive years. To solve this problem, we need to forecast the asset price in a given time horizon based on the historical data. This means that it is necessary to obtain information about the distribution of the stock price and its log return from a limited number of historical data. To tackle this problem, researchers try to enlarge the size of data set to analyse the distribution of statistics of interest by resampling. The resampled data set should maintain the property of the original distribution, from which a sample is obtained.

To achieve this purpose, bootstrap and other resampling methods are introduced to solve this problem. Since the previous tests [16, 23] show that the stock price has serial correlation, block bootstrap resampling is widely used to enlarge the data set. In financial research, we often conduct distribution analysis to estimate the asset log returns. Thus, the performance of block bootstrap resampling in estimating the distribution of log return, especially the total log return in a given time period, is considered. In this research, we discuss the performance of block bootstrap resampling for estimating the bootstrapped monthly return and the total return in a given time horizon. The stationary block bootstrap resampling is the main method used for resampling.

In this research, we use a Geometric Brownian Motion (GBM) and a Customized AR(1) model first introduced in [25] as the stock price model. For the Customized AR(1) model, we derive the continuous form and use it for further analysis. We follow the similar bootstrap procedure in [7] to conduct block bootstrap resampling from a sample path of 100-year monthly log returns obtained from Monte Carlo simulation. We use the path to generate 30-year monthly log return paths. We then compare the statistics of monthly and total log return (summation of each monthly return path) with the theoretical values derived from the continuous form of the stock price model. For the Customized AR(1) model, before the analysis of bootstrapped distribution, we compare the bias of stationary block bootstrap resampling in estimating the variance ratio $VR(k)$ (ratio of variance between m-period and 1-period variance), as is done in a discrete form in [5, 6].

For the monthly log return, we find that the block bootstrap resampling performs well in predicting the distribution statistics (e.g. mean and variance) for these two models, no matter how large the block size is.

For the total log return, which is the summation of monthly returns for each path, we find that for Geometric Brownian Motion, the bootstrapped estimation with block size of one month outperforms other block sizes. This is because the log return of GBM follows a standard Brownian motion. Thus, there exists no correlation between monthly log returns. When using a large block size, the resampling triggers high similarity among resampling paths, and thus the distribution estimation of total log return becomes poor.

For the Customized AR(1) model, before the analysis of the total return distribution, we test the bootstrap performance of the variance ratio estimation in the continuous form with that in the discrete form in [5, 6]. It is found that the results are similar, and the continuous form even provides a smaller estimation bias when the volatility ratio is larger than one. We further analyse bootstrap resampled results with different model parameters and different volatility ratios of the noise term in the permanent and transitory component in the model. For the bootstrap resampling for the total log return, we choose block sizes from one month to 150 months. We find the bootstrapped distribution to be significantly poor with chosen block sizes. This is due to the high autocorrelation between log returns for the Customized AR(1) model. Unlike the GBM, the small block size destroys the serial dependence between monthly return along the sample path. In this case, the method does not perform well in the estimation of distribution.

Based on the definition in [20], since we use monthly returns as a sample path, the distribution statistics (e.g. mean $E(X)$, variance $Var(X)$ and total log return) for monthly return is the level-1 parameter, while distribution statistics of the level-1 parameter is called the level-2 parameter (e.g., statistics of total log return). The definition of level-1 and higher-level parameters is included in Chapter 2 in detail. In this research, we find that the block bootstrap resampling results has trivial differences in estimating level-1 parameters, no matter how large the correlation within the sample paths. However, for estimating level-2 parameters, there exists estimation bias among different block sizes. To mitigate this bias, as [18, 14] mention, an optimal block size should be selected during the block bootstrap resampling for estimation accuracy.

The rest of the presentation is organized as follows. In Chapter 2, the background information for block bootstrap resampling and the estimation bias of block bootstrap resampling in theoretical analysis are briefly summarized. Chapter 3 derives the formations for log return and the continuous form of two stock price models. Chapter 4 includes the Monte Carlo algorithms for monthly and total log return. Chapter 5 describes the bootstrap resampling procedure with a mean-shifting method for bootstrapped distribution, and Chapter 6 provides the bootstrapped analysis for two stock price models. Chapter 7 concludes the research and propose the further work on this topic.

Chapter 2

Background

In this chapter, we introduce the background information about bootstrap resampling. First we review the development of bootstrap resampling and some block bootstrap resampling methods which are widely used at the present time. Then we provide a brief introduction about the estimation bias of block bootstrap resampling.

2.1 Bootstrap resampling

According to [20], the bootstrap is a computer-intensive resampling method for solving statistical estimation problems. The sampling is done with replacement. In addition, there is no prerequisites of samples and models. The standard bootstrap resampling is first introduced in [9]. It is designed for samples that are independent and identically distributed (i.i.d). Every data point is selected randomly from the sample path with equal probability. Bootstrap resampling has been proved to be robust and consistent [2, 15, 31] for mean and variance estimation. It has been widely used for many statistical inference problems. Recently, it has been further developed and classified depending on whether the sampling path is a set of independent or dependent data.

2.1.1 Block bootstrap resampling

As mentioned before, the standard bootstrap resampling in [9] is for i.i.d samples. However, the data from the financial market typically has serial correlation (for example, [16, 23] determined that asset prices do not follow a pure random walk). In order to maintain the correlation in resampling data set, block bootstrap resampling is proposed as an extension of standard bootstrap resampling. Instead of resampling every data point independently, the algorithm introduces block size b and select the current and the following $b-1$ data points simultaneously as a "block", which helps to maintain the serial correlation between the original sample.

Assume $X = \{X_1, X_2, \ldots, X_N\}$ is the sample path. Before the block bootstrap resampling, a block size b is determined. In addition, the number of "blocks" l which is required

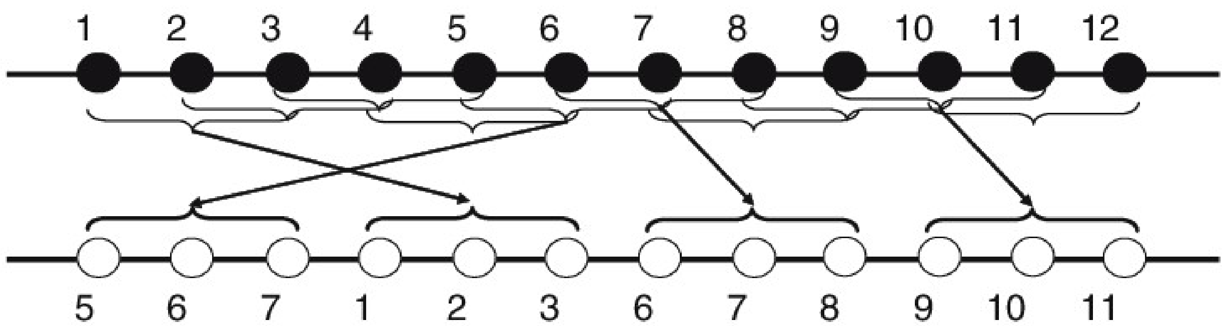

Figure 2.1: Process of Moving Block bootstrap resampling from [13] (up: sample path, down: resampling path)

in a sample path is determined. We can treat the standard bootstrap resampling as block bootstrap resampling with $b = 1$ and $l = N$.

There are mainly three block bootstrap methods widely used, which are the moving block bootstrap resampling (MBB), the non-overlapping block bootstrap resampling (NBB) and the stationary block bootstrap resampling (SB). The moving block bootstrap resampling was independently developed in [18, 22]. As shown in Figure 2.1), it allows overlapping in sampling blocks and each block is represented as $B_i = \{X_i, X_{i+1}, \ldots, X_{i+b-1}\}$ and the number of blocks in sample path $l = N - b + 1$. The non-overlapping block bootstrap resampling was introduced in [4]. The blocks in sample path are not overlapping, i.e., $B_i = \{X_{(i-1)\times b+1}, X_{(i-1)\times b+1}, \ldots, X_{i\times b}\}$ and the number of blocks $l = N/b$. These two methods share the property that the block size b is fixed for each block (In Figure 2.1, the block size is fixed at 3). However, to eliminate the bias of using a fixed block size, [28] introduced the stationary block bootstrap resampling. In this method, the block size used in resampling varies from block to block, and the number of blocks l changes with the resampling path.

Following the definition in [20], suppose the sample path $X = \{X_1, X_2, \ldots, X_N\}$ has a joint distribution J. Let $\theta \equiv \theta(J)$ be a point estimator of X, which is also the parameter of interest that is based on the joint distribution J. We call θ as the level-1 parameter (e.g., mean $E(X)$ and variance $Var(X)$). Statistics which depend on the sampling distribution of θ are called level-2 parameters (e.g., mean square error $MSE(E(X)), Var(E(X))$). Based on this definition, we have level-3 and higher-level parameters. The bootstrap resampling is considered as a suitable method for solving the statistical inference problem for level-2 parameters.

2.1.2 Estimation bias of block bootstrap resampling

Although block bootstrap resampling helps to increase the size of data sets without any assumptions, as a resampling method with random selection, there exists estimation bias which should be analysed.

Theoretically, it is proved in [19] that the estimation bias for bootstrap resampled variance of resampling data is proportional to $1/(nb)$ where n is the resampling size of data and b is the block size, where those three resampling methods mentioned above (MBB, NBB and SB) share the same magnitude of estimation error. In practice, according to [5, 6], the estimation bias of variance ratio using moving and stationary block bootstrap resampling is significant and increasing with the increase of the time horizon and the volatility ratio. In [20], several bootstrap resampled results, such as the sampling distribution of sample mean and covariance of bootstrap resampled data, show that the estimation bias varies among different block sizes.

According to [1], a too-large and too-small block size will decrease the accuracy of bootstrap resampled estimation, so to mitigate this effect, researchers have developed methods to determine the optimal block size for block bootstrap resampling. In [27, 29], an automatic selection of block size is introduced for minimizing the long-run variance of the resampling path. Following up on this work, [1, 5, 6] apply a data-driven method for selecting the optimal block size that minimizes the estimation bias of statistics with interest. Both these two methods are widely applied in the research related to block bootstrap resampling.

Chapter 3

Models for log return and stock price

In this chapter, we set up the definition of log return and describe models based on the assumption that the stock price follows a Geometric Brownian Motion (GBM) or a Customized AR(1) model introduced in [25]. For the Customized AR(1) model, we convert the discrete form into a continuous form with a combination of a Standard Brownian Motion and an Ornstein-Uhlenbeck Process (OU Process). In the rest of this research, we use the continuous form of both models for further analysis and computation. Finally, we derive the statistics of log return for a given time horizon.

3.1 Log return

Let the initial time $t_0 = 0$ and time horizon $t_N = T$. Let $\Delta t = t_n - t_{n-1}$ represents the time step, $N = T/\Delta t$ is the total number of time steps within T, $n \in [1, N]$. S_{t_n} denotes the spot asset price at time t_n and $R(t_n)$ is the log return in the period $[t_{n-1}, t_n)$, $n \in [1, T]$. Then $R(t_n)$ is

$$R(t_n) = log\Big(\frac{S_{t_n}}{S_{t_{n-1}}}\Big). \tag{3.1a}$$

The total log return for the time horizon T, R_T, is the summation of log return for each single period, i.e.,

$$R_T = \sum_{n=1}^{N} R(t_n) = log\Big(\frac{S_{t_N}}{S_{t_0}}\Big). \tag{3.1b}$$

If we define X_{t_n} as the log of stock price S_{t_n}, i.e., $X_{t_n} = log(S_{t_n})$. Then

$$R_T = X_{t_N} - X_{t_0}. \tag{3.1c}$$

3.2 Geometric Brownian Motion

In this section, suppose the stock price S_{t_n} follows GBM, and we derive the formations of X_{t_n} and R_T, and calculate the statistics of R_T.

3.2.1 Continuous form of X_{t_n}

If we assume that stock price S_t follows a GBM, the stochastic differential equation for S_t is

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{3.2}$$

where μ and σ are the mean and volatility of the asset price S_t, and dW can be written as

$$dW_t = \psi\sqrt{dt} = \mathcal{N}(0,1)\sqrt{\Delta t}, \tag{3.3}$$

where ψ is a random variable following a standard normal distribution, i.e., $\psi \sim \mathcal{N}(0,\, 1)$.

Based on Ito's Lemma, the continuous form of $X_t = log(S_t)$ is

$$dX_t = (\mu - \frac{\sigma^2}{2})dt + \sigma dW_t, \tag{3.4}$$

Hence the discretization of X_{t_n} can be set as

$$X_{t_n} = X_{t_{n-1}} + (\mu - \frac{\sigma^2}{2})\Delta t + \sigma\sqrt{\Delta t}\,\mathcal{N}(0,1), \tag{3.5}$$

3.2.2 Statistics for R_T

Based on (3.5), we derive

$$R(t_n) = (\mu - \frac{\sigma^2}{2})\Delta t + \sigma\sqrt{\Delta t}\,\mathcal{N}(0,1) = (\mu - \frac{\sigma^2}{2})\Delta t + \mathcal{N}(0,\sigma^2\Delta t) \tag{3.6}$$

then from (3.1b), we have

$$R_T = \sum_{n=1}^{N} R(t_n) = (\mu - \frac{\sigma^2}{2})\,T + \mathcal{N}(0,\,\sigma^2 T), \tag{3.7}$$

then the statistics for R_T are easily determined.

For mean $E(R_T)$, variance $Var(R_T)$ and standard deviation $std(R_T)$:

$$E(R_T) = (\mu - \frac{\sigma^2}{2})\,T, \quad Var(R_T) = \sigma^2 T, \quad std(R_T) = \sigma\sqrt{T}, \tag{3.8a}$$

For skewness $skewness(R_T)$:

$$\begin{aligned} skewness(R_T) &= \frac{E\Big(\big(R_T - E(R_T)\big)^3\Big)}{(std(R_T))^3} \\ &= \frac{E\Big(\big(\sigma\sqrt{T}\,\mathcal{N}(0,\,1)\big)^3\Big)}{(std(R_T))^3} \\ &= \frac{(\sigma\sqrt{T})^3 E\Big(\big(\mathcal{N}(0,\,1)\big)^3\Big)}{(std(R_T))^3} \\ &= 0 \end{aligned} \tag{3.8b}$$

For kurtosis $kurtosis(R_T)$:

$$
\begin{aligned}
kurtosis(R_T) &= \frac{E\Big(\big(R_T - E(R_T)\big)^4\Big)}{(std(R_T))^4} \\
&= \frac{E\Big(\big(\sigma\sqrt{T}\,\mathcal{N}(0,\,1)\big)^4\Big)}{(std(R_T))^4} \\
&= \frac{(\sigma\sqrt{T})^4 E\Big(\big(\mathcal{N}(0,\,1)\big)^4\Big)}{(std(R_T))^4} \\
&= \frac{3\;(\sigma\sqrt{T})^4}{(std(R_T))^4} \\
&= 3.
\end{aligned}
\tag{3.8c}
$$

3.3 Customized AR(1) model

In this section, assuming the asset price follows a Customized AR(1) discrete model, we derive the continuous form of X_t, and the statistics of R_T based on X_t. Also, we introduce two indicators related to the serial dependence for the continuous model of the Customized AR(1) for further analysis.

3.3.1 Discrete form

The customized AR(1) model was first introduced in [25] and widely applied in research as a simple model for an asset price. According to [6, 10], the log of stock price X_{t_n} could be generally divided into two components: a permanent component D_{t_n} which is a random walk representing the fundamental increment of the asset price, and a transitory component V_{t_n} which is a zero-mean stationary process as an AR(1) model representing the volatility from D_{t_n}. Specifically,

$$
\begin{aligned}
X_{t_n} &= log(S_{t_n}), \\
X_{t_n} &= D_{t_n} + V_{t_n}, \\
D_{t_n} &= D_{t_{n-1}} + \mu\Delta t + \epsilon_{t_n}, \\
V_{t_n} &= \phi V_{t_{n-1}} + w_{t_n}
\end{aligned}
\tag{3.9}
$$

where μ is the expected drift, ϕ is the model parameter where $\phi < 1$, and ϵ_{t_n}, w_{t_n} are white noises where $cov(\epsilon_{t_n}, w_{t_n}) = 0$.

3.3.2 Continuous form

In this section, we introduce two continuous processes, Standard Brownian Motion and Ornstein-Uhlenbeck Process, and transfer the customized AR(1) model into its continuous form based on these two processes.

Standard Brownian Motion

We assume that Y_{t_n} follows a Standard Brownian Motion, and the stochastic differential equation of Y_t is

$$dY_t = \mu_1 dt + \sigma_1 dW_t, \tag{3.10}$$

where μ_1, σ_1 are constants and dW_t follows the process

$$dW_t = \psi_1 \sqrt{dt} \tag{3.11}$$

where ψ_1 follows a standard normal distribution, $\psi_1 \sim \mathcal{N}_1(0, 1)$. The discretization of Y_{t_n} is

$$Y_{t_n} = Y_{t_{n-1}} + \mu_1 \Delta t + \sigma_1 \sqrt{\Delta t}\, \mathcal{N}_1(0, 1), \tag{3.12}$$

where $\Delta t > 0, \sigma_1 > 0$, μ_1 is a constant.

Ornstein-Uhlenbeck Process (OU Process)

We assume Z_t follows an Ornstein-Uhlenbeck process, which is a stationary process. The stochastic differential equation of Z_t is

$$dZ_t = \theta_2(\mu_2 - Z_t)dt + \sigma_2 dW_t, \tag{3.13}$$

where $\theta_2 > 0, \sigma_2 > 0$, μ_2 is a constant, and dW_t can be written as

$$dW_t = \psi_2 \sqrt{dt}, \tag{3.14}$$

where ψ_2 follows standard normal distribution, i.e., $\psi_2 \sim \mathcal{N}_2(0, 1)$. The discretization of Z_{t_n} is

$$Z_{t_n} = \theta_2 \mu_2 \Delta t + (1 - \theta_2 \Delta t) Z_{t_{n-1}} + \sigma_2 \sqrt{\Delta t}\, \mathcal{N}_2(0, 1), \tag{3.15}$$

where $\theta_2 > 0, \Delta t > 0, \sigma_2 > 0$, and μ_2 is a constant.

Continuous form of Customized AR(1) discrete model (3.9)

Since D_{t_n} follows a random walk and V_{t_n} follows an AR(1) process, we show in the following that the continuous-time analogue of D_{t_n} and V_{t_n} are a Standard Brownian Motion and an Ornstein-Uhlenbeck process.

For D_{t_n}, comparing (3.9) and (3.12), it is found that if we set $\mu_1 = \mu$, $\epsilon_{t_n} \sim \mathcal{N}_1(0,\, {\sigma_1}^2 \Delta t)$, then we can write D_{t_n} as the discretization of Standard Brownian Motion that

$$D_{t_n} = D_{t_{n-1}} + \mu \Delta t + \sigma_1 \sqrt{\Delta t} \mathcal{N}_1(0,\, 1). \tag{3.16a}$$

The continuous form of D_{t_n} can be expressed as

$$dD_t = \mu dt + \psi_1 \sigma_1 \sqrt{dt}. \tag{3.16b}$$

where μ is a constant, $\psi_1 \sim \mathcal{N}_1(0,\, 1)$.

For V_{t_n}, comparing (3.9) and (3.15), it is found that $\theta_2 \mu \Delta t = 0,\ 1 - \theta_2 \Delta = \phi,\ w_{t_n} \sim \mathcal{N}_2(0,\, {\sigma_2}^2 \Delta t)$. To maintain Z_{t_n} is a stationary process, we have $\phi < 1$, so it is derived that $\mu_2 = 0,\ \theta_2 \Delta t = 1 - \phi,\ w_{t_n} \sim \mathcal{N}_2(0,\, {\sigma_2}^2 \Delta t)$. Then we can write V_{t_n} as the discretization of an OU Process that

$$V_{t_n} = (1 - \theta_2 \Delta t) V_{t_{n-1}} + \sigma_2 \sqrt{\Delta t} \mathcal{N}_2(0,\, 1), \tag{3.17a}$$

so the continuous form of V_{t_n} is expressed as

$$dV_t = -\theta_2 V_t dt + \psi_2 \sigma_2 \sqrt{dt}. \tag{3.17b}$$

where θ_2 is a constant, $\psi_2 \sim \mathcal{N}_2(0,\, 1)$.

The Customized AR(1) model (3.9) can be regarded as the following continuous model:

$$dX_t = dD_t + dV_t \tag{3.18a}$$

where

$$dD_t = \mu dt + \psi_1 \sigma_1 \sqrt{dt} \tag{3.18b}$$

$$dV_t = -\theta_2 V_t dt + \psi_2 \sigma_2 \sqrt{dt} \tag{3.18c}$$

where μ is a constant, $\sigma_1 > 0, \sigma_2 > 0,\ \psi_1, \psi_2$ follows standard normal distribution where $cov(\psi_1, \psi_2) = 0$. θ_2 is a model parameter in the continuous form. In addition, the following relationship holds

$$\theta_2 \Delta t = 1 - \phi. \tag{3.18d}$$

The discretization of the continuous form of customized AR(1) model is (3.16a) and (3.17a).

3.3.3 Statistics of R_T

In this section, we derive the mean and variance of log return R_T, assuming the continuous model (3.18). First, we would like to derive the statistics for X_t. Since in (3.18), $cov(\psi_1, \psi_2) = 0$, $cov(dD_t, dV_t) = 0$, we have

$$\begin{aligned} E(dX_t) &= E(dD_t) + E(dV_t), \\ Var(dX_t) &= Var(dD_t) + Var(dV_t). \end{aligned} \tag{3.19}$$

We derive the mean and variance of dD_t and dV_t.

For dD_t, based on (3.18b),

$$E(dD_t) = \mu dt, \quad Var(dD_t) = \sigma_1^2 dt. \tag{3.20a}$$

In addition,

$$\begin{aligned} E(D_{t_N} - D_{t_0}) &= \mu T, \\ Var(D_{t_N} - D_{t_0}) &= \sigma_1^2 T. \end{aligned} \tag{3.20b}$$

For dV_t, let $\bar{V}_{t_n} = E(V_{t_n}), V_{t_0} = 0, \bar{V}_{t_0} = E(V_{t_0}) = 0$,

$$\begin{aligned} d\bar{V}_t = dE(V_t) = E(dV_t) = E(-\theta_2 V_t dt + \psi_2 \sigma_2 \sqrt{dt}) &= -\theta_2 \bar{V}_t dt, \\ E(V_{t_N}) = \bar{V}_{t_N} = \bar{V}_{t_0} exp(-\theta_2 t) &= 0, \end{aligned} \tag{3.21a}$$

Let $G(V,t) = V^2$, then $G_V = 2V,\ G_{VV} = 2,\ G_t = 0$ (here G_t is the partial differential of G with respect to t), then according to Ito's lemma and (3.18c), let $a = -\theta_2 V_2,\ b = \sigma_2$,

$$\begin{aligned} dG &= (aG_V + G_t + \frac{b^2}{2} G_{VV})\ dt + b\ G_V dW_t \\ &= (-2\theta_2 V^2 + \sigma_2^2) dt + 2\sigma_2 V dW_t. \end{aligned} \tag{3.21b}$$

Let $\bar{G} = E(G) = E(V^2)$,

$$\begin{aligned} d\bar{G} &= dE(V^2) = E(dV^2) \\ &= E\big((-2\theta_2 V^2 + \sigma_2^2)\ dt + 2\sigma_2 V dW_t\big) \\ &= \sigma_2^2 dt - 2\theta_2 E(V^2 dt) \\ &= \sigma_2^2 dt - 2\theta_2 \bar{G} dt, \end{aligned} \tag{3.21c}$$

which is a first-order ordinary differential equation. Let $y = \bar{G}$, then

$$\frac{dy}{dt} + 2\theta_2 y = \sigma_2^2. \tag{3.21d}$$

We multiply by an integrating factor $e^{\int_0^t 2\theta_2 ds} = e^{2\theta_2 t}$ in (3.21d),

$$\begin{aligned} e^{2\theta_2 t} \frac{dy}{dt} + 2\theta_2 e^{2\theta_2 t} y &= e^{2\theta_2 t} \sigma_2^2 \\ d(e^{2\theta_2 t} y) &= e^{2\theta_2 t} \sigma_2^2 \\ e^{2\theta_2 t} y &= \int_0^t e^{2\theta_2 t} \sigma_2^2 ds = \frac{\sigma_2^2 (e^{2\theta_2 t} - 1)}{2\theta_2} \\ y &= \frac{\sigma_2^2 (1 - e^{-2\theta_2 t})}{2\theta_2}, \end{aligned} \tag{3.21e}$$

so we have

$$E(V_{t_N}^2) = \frac{\sigma_2^2 (1 - e^{-2\theta_2 T})}{2\theta_2}, \tag{3.21f}$$

so the variance of V_{t_N} is derived as

$$Var(V_{t_N}) = E(V_{t_N}^2) - \big(E(V_{t_N})\big)^2 = \frac{\sigma_2^2 (1 - e^{-2\theta_2 T})}{2\theta_2}, \tag{3.21g}$$

Hence, based on (3.1c), (3.19), (3.20b), (3.21a), (3.21) and $V_{t_0} = 0$,

$$\begin{aligned} E(R_T) &= E(X_{t_N} - X_{t_0}) \\ &= E(D_{t_N} - D_{t_0}) + E(V_{t_N} - V_{t_0}) = \mu T, \\ Var(R_T) &= Var(X_{t_N} - X_{t_0}) \\ &= Var(D_{t_N} - D_{t_0}) + Var(V_{t_N} - V_{t_0}) \\ &= Var(D_{t_N} - D_{t_0}) + Var(V_{t_N}) \\ &= \sigma_1^2 T + \frac{\sigma_2^2(1 - e^{-2\theta_2 T})}{2\theta_2}. \end{aligned} \tag{3.22a}$$

3.3.4 Serial dependence indicators

Since the log return defined by the Customized AR(1) model is serially dependent, we introduce two indicators, the variance ratio $VR(k)$ and the first order auto-correlation $AC1(k)$, to quantify the correlation in terms of the model parameter θ_2 and the volatility ratio for white noises in D_{t_n} and V_{t_n}.

Variance ratio

The variance ratio was first introduced in [23] which tests whether a sample series follows random walk or not. As in [6], it is formulated as

$$VR(k) = \frac{Var(R_{i,i+k})}{k\ Var(R(t_i))}, \tag{3.23}$$

where $R_{i,i+k} = \sum_{j=1}^{k} R(t_{i+j-1})$, $Var(R(t_i))$ and $Var(R_{i,i+k})$ are the $1-$ and $k-$period variance of the sample series. If the log return $R(t_i)$ is a series following a random walk, which means that there is no correlation between $R(t_i)$ and $R(t_{i+k})$, then $Var(R_{i,i+k}) = Var(R(t_i) + \cdots + R(t_{i+k-1})) = kVar(R(t_i))$. Hence, $VR(k) = 1$. Otherwise, the value will diverge from one.

Based on [6], with respect to the continuous model parameter θ_2 of the Customized AR(1) model, the variance ratio $VR(k)$ is derived as

$$VR(k) = \frac{k(1 - (1-\theta_2)^2)\ r^2 + 2(1 - (1-\theta_2)^m)}{k(1 - (1-\theta_2)^2)\ r^2 + 2k\theta_2}, \tag{3.24}$$

where $r = \sigma_1/\sigma_2$ is the volatility ratio for white noises in D_{t_n} and V_{t_n}. Figure 3.1 shows how $VR(k)$ decays as the time period, the volatility ratio and the model parameter increases. It is also found that the decreasing speed of $VR(k)$ slows down with the increases of these variables. This means when the white noise in D_{t_n} dominates the variation of the whole model, the auto-correlation between samples decreases faster in the long run. At the same time, $VR(k)$ decays slower which is shown in Figure 3.1.

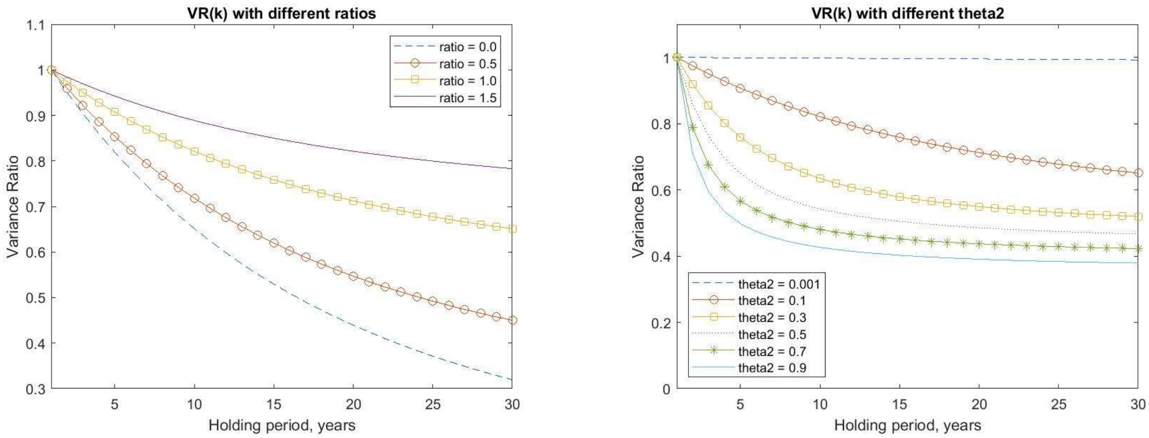

Figure 3.1: Variance ratio $VR(k)$ for different volatility ratios and different model parameters (Left: $\theta_2 = 0.1, r = 0.0,\ 0.5,\ 1.0,\ 1.5$; Right: $r = 1, \theta_2 = 0.001,\ 0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9$)

First order auto-correlation

The first order auto-correlation $AC1(k)$ is designed to measure the correlation between compounded $k-$period log return. In [5], it is formulated as

$$AC1(k) = \frac{Cov(X_{i,i+k}, X_{i-k,i})}{\sqrt{Var(X_{i,i+k})Var(X_{i-k,i})}}, \tag{3.25}$$

and with respect to θ_2, it is derived as

$$AC1(k) = \frac{(1-\theta_2)^k(2-(1-\theta_2)^k)-1}{k(1-(1-\theta_2)^2)\ r^2 + 2(1-(1-\theta_2)^k)}, \tag{3.26}$$

where $r = \sigma_1/\sigma_2$ is the volatility ratio for the white noise in D_{t_n} and V_{t_n}. Figure 3.2 shows the first order auto-correlation of $k-$period return with the change of θ_2 and ratio r. It is found that the log returns of the Customized AR(1) model are generally negatively correlated, and with the increase of r and θ_2, the absolute value of auto-correlation for $k-$period return generally decreases. This trend is explained from the continuous model (3.18) that under this two conditions, the variance of the model mainly stems from the white noise in D_{t_n} and the model tends to a random walk, and the absolute value of auto-correlation converges to zero. This result also matches the proof in [10] that the auto-correlation $AC1(k)$ of the customized AR(1) model is negative, and if θ_2 tends to zero, it approaches -0.5 for large values of k.

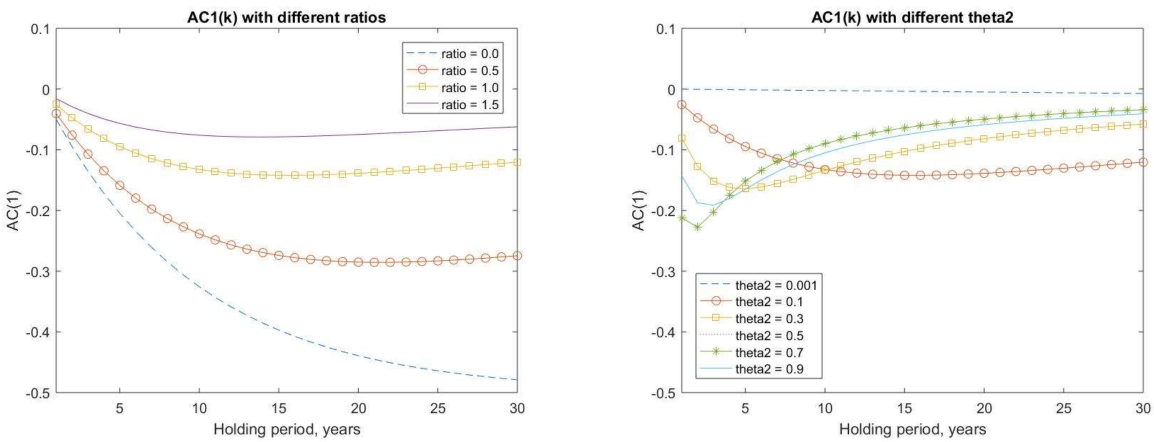

Figure 3.2: First order auto-correlation for different volatility ratios and different model parameters (Left: $\theta_2 = 0.1, r = 0.0$, 0.5, 1.0, 1.5; Right: $r = 1, \theta_2 =$0.001, 0.1, 0.3, 0.5, 0.7, 0.9)

Chapter 4

Algorithms

In this chapter, we introduce the Monte Carlo simulation algorithms for the total log return and the sample path in a given time horizon for each stock price model. To verify the correctness of those algorithms, we not only compare the simulation results with the theoretical values derived in the previous chapter, but also among different simulation times for the total log return, and different time horizons for the sample path of monthly return. For the total log return, it is expected that the distribution statistics of the total log return converges to the theoretical value with an increasing number of simulations. For the sample path monthly log return, when the time horizons increases, the sample size increases. In this case, the sampled distribution statistics will experience the same change as simply increasing of number of simulations. In this testing, we focus on the estimates of mean and variance (Var).

4.1 Algorithms

4.1.1 Simulation for R_T under the Geometric Brownian Motion

According to (3.5), the simulation for R_T under the Geometric Brownian Motion is described in Algorithm 1. In simulation, we set the time step to be sufficiently small to imitate the log of stock price in the continuous model (3.4). Thus, we set $\Delta t = 1/252$ year.

4.1.2 Simulation for R_T under the Customized AR(1) model

According to (3.1c) and (3.9), the simulation for R_T under the Customized AR(1) is described in Algorithm 2. Following the same setting in the previous algorithm, we set $\Delta t = 1/252$ year.

Algorithm 1 Log return under the Geometric Brownian Motion

Input: Time horizon T years; Initial asset price S_{t_0}; Mean rate of return μ; Volatility σ

Initialize: Time step $\Delta t = 1/252$ year; Number of steps $N = T/\Delta t$; $X_{t_0} = log(S_{t_0})$

Output: R_T

```
for i = 1, 2, ..., N do
    X_{t_i} = X_{t_{i-1}} + (μ − σ²/2)Δt + σ√Δt 𝒩(0, 1)
end for
R_T = X_{t_N} − X_{t_0}
```

$$\textbf{for } i = 1, 2, ..., N \textbf{ do}$$
$$\quad X_{t_i} = X_{t_{i-1}} + (\mu - \tfrac{\sigma^2}{2})\Delta t + \sigma\sqrt{\Delta t}\,\mathcal{N}(0,\,1)$$
$$\textbf{end for}$$
$$R_T = X_{t_N} - X_{t_0}$$

Algorithm 2 Log return of Customized AR(1) model

Input: Time horizon T years; Mean rate of return μ; Volatility of ϵ_{t_n} σ_1; Volatility of w_{t_n} σ_2; Model parameter θ_2

Initialize: Time step $\Delta t = 1/252$ year; Number of steps $N = T/\Delta t$; $D_{t_0} = 1$; $V_{t_0} = 0$

Output: R_T

$$\textbf{for } i = 1, 2, ..., N \textbf{ do}$$
$$\quad D_{t_i} = D_{t_{i-1}} + \mu\Delta t + \sigma_1\sqrt{\Delta t}\,\mathcal{N}_1(0,\,1)$$
$$\quad V_{t_i} = (1 - \theta_2\Delta t)V_{t_{i-1}} + \sigma_2\sqrt{\Delta t}\,\mathcal{N}_2(0,\,1)$$
$$\textbf{end for}$$
$$R_T = X_{t_N} - X_{t_0} = (D_{t_N} + V_{t_N}) - 1$$

4.2 Simulation results

4.2.1 Simulation results for R_T

Tables 4.1 and 4.2 present the parameter for the simulation of R_T under the Geometric Brownian Motion and the Customized AR(1) model. We set $T = 30$ years and $N_{sim} = 10000, 50000, 100000, 200000$, and we simulate the total log return R_T based on Algorithm 1 and 2. Table 4.3 and 4.4 show that with the number of simulations increases, the statistics converges to the theoretical values, which is consistent with our expected results.

S_0	μ	σ
100	0.01	0.2

Table 4.1: Inputs for simulation of R_T of the Geometric Brownian Motion

μ	σ_1	σ_2	θ_2
0.01	0.2	0.2	0.1

Table 4.2: Inputs for simulation of R_T of the Customized AR(1)

Statistics	Theoretical	Number of simulations			
		10,000	50,000	100,000	200,000
Mean	-0.300000	-0.308797	-0.302403	-0.297940	-0.302028
Var	1.200000	1.167504	1.196725	1.206749	1.199948
SD	1.095445	1.080511	1.093949	1.098521	1.095422

Table 4.3: Statistics for R_T of the Geometric Brownian Motion with number of simulations $N_{sim} = 10000, 50000, 100000, 200000$.

Statistics	Theoretical	Number of simulations			
		10,000	50,000	100,000	200,000
Mean	0.300000	0.321020	0.296936	0.296795	0.302166
Var	1.399504	1.443520	1.402751	1.402336	1.399742
SD	1.183006	1.201466	1.184378	1.184203	1.183107

Table 4.4: Statistics for R_T of the Customized AR(1) model with number of simulations $N_{sim} = 10000, 50000, 100000, 200000$.

4.2.2 Sample path simulations

In this research, we denote the time horizon of a sample path as $T^{'}$, N_1 represents the number of months in the time horizon T', i.e., $N_1 = 12T'$. Then, a sample path of monthly returns using block bootstrap resampling as

$$R_{sample} = \{R_{sample,1}, R_{sample,2}, ..., R_{sample,N_1}\},$$

We use the same inputs in Table 4.1 and 4.2, and set the number of simulations $N_{sim} = 200,000$, time horizon $T' = 100, 200, 500, 1000, 2000$, and sample the Monte Carlo simulation monthly. Table 4.5 and 4.6 show that, with the increase of time horizon, the simulated estimations converge to the theoretical values.

Statistics	Theoretical	Time horizon (years)				
		100	200	500	1000	2000
Mean	-0.0008	-0.0028	-0.0024	-0.0013	-0.0007	-0.0009
Var	0.0033	0.0034	0.0033	0.0033	0.0034	0.0033
SD	0.0577	0.0580	0.0572	0.0573	0.0580	0.0579

Table 4.5: Statistics for sampled monthly log return under the Geometric Brownian Motion with time horizon $T' = 100, 200, 500, 1000, 2000$ years.

Statistics	Theoretical	Time horizon (years)				
		100	200	500	1000	2000
Mean	0.0008	0.0022	0.0004	0.0013	0.0005	0.0009
Var	0.0066	0.0069	0.0065	0.0066	0.0066	0.0066
SD	0.0815	0.0832	0.0809	0.0811	0.0815	0.0815

Table 4.6: Statistics for sampled monthly log return under the Customized AR(1) model with time horizon $T' = 100, 200, 500, 1000, 2000$ years.

Chapter 5

Block bootstrap resampling

In the financial market, the historical data of an asset price for a long time horizon is insufficient for further research, especially when we desire to study its distribution where a large sample size is a fundamental requirement for accurate estimations. To solve this problem, we try to enlarge the data set for distribution analysis. For example, bootstrap resampling is widely used. There is good evidence that stock prices have some serial dependence, i.e., [16, 23]. To retain this dependence in the resampled data set, block bootstrap resampling, especially stationary block bootstrap resampling where the actual block size varies for each block, has been used, e.g. [12, 7, 11]. In this research, we bootstrap resample the log return.

Before implementing bootstrap resampling, we use Monte Carlo simulation to obtain the distribution of total log return R_T for a given time horizon T. Then we conduct block bootstrap resampling for the bootstrap resampled distribution of total log return $R_{bootstrap}$ with the same horizon.

The process of resampling is as follows. First, for each block bootstrap resampling distribution, we generate a single sample path of monthly return R_{sample} using Monte Carlo simulation where the time horizon $T^{'}$ is larger than the given investment horizon T. Those sample paths are determined using the optimal block size for long-run variance estimation. After that, we conduct stationary block bootstrap resampling based on these sample paths, sum up the monthly return in each resampling path for the bootstrap resampled total log return $R_{bootstrap}$. Finally, we investigate differences between the true and the bootstrap resampled distribution.

5.1 Stationary block bootstrapping

As mentioned before, stationary block bootstrapping was first introduced in [28] to avoid the impact of using a fixed block size. Let $p = \frac{1}{b_{exp}}$ where b_{exp} is the expected block size, and the actual block size b_{ac} used for each block bootstrap resampling is generated by a geometric distribution that

$$Pr(Y = b_{ac}) = (1 - p)^{b_{ac}-1}p. \tag{5.1}$$

We choose different expected block sizes b_{exp}. We implement the process of bootstrapping in [27, 29] to resample each block and concatenate them to form a full path of monthly log returns for the given time horizon T. The algorithm is sketched in Algorithm 3.

Algorithm 3 Stationary block bootstrapping

Input: sample path R_{sample}, length of bootstrap resampled sample N, expected block size b_{exp}

Initialize: n = number of period (monthly) returns in R_{sample}, sub_block_total = n, total_sample = 1, actual block size $b_{ac} = 0$

Output: $R_{bootstrap,month}$

```
while total_sample ≤ N do
    if sub_block_total > b_ac then
        Generate a random starting index in sample path to bootstrap, and round the
        index to the nearest integer i.e. index = round(1 + rand(1,1) × (n − 1))
        Generate actual block size from the geometric distribution, i.e. b_ac = geo(b_exp)
        sub_block_total = 1
    end if
    if index > n then
        index = index - n
    end if
    R_bootstrap(t) = R_sample(b_ac)
    index = index + 1
    sub_block_total = sub_block_total + 1
    total_sample = total_sample + 1
end while
```

After bootstrap resampling N_{sim} times, we obtain the bootstrap resampled monthly return $R_{bootstrap,month}$ and sum it up for each path, i.e.,

$$R_{bootstrap,T}(i) = \sum_{j=1}^{N} R_{bootstrap,month}(i,j), \;\; i \in [1, N_{sim}] \tag{5.2}$$

to obtain bootstrap resampled log return $R_{bootstrap,T}$ for the given time horizon T.

5.2 Distribution shifting

When studying the distribution of R_{sample} of the two models discussed in the previous chapter, it is found that when performing Monte Carlo simulation to the model directly, the estimation of variance converges quickly to the true value, while the convergence of mean takes place only when the time horizon of the sample path T' is sufficiently large. Table 5.1 shows the mean estimation of total log return for different sample paths under the Geometric Brownian Motion. It is found that the block bootstrap is less satisfactory in approximating the mean of total log return for all block sizes. What is more, for all these examples, the gap between the theoretical and resampled value is significant and path dependent. The reason is that we only use one sample path with a limited time horizon in each block bootstrap resampling, where the monthly return along the sample path is randomly simulated. This is not supposing, since the total log return only depends on the initial and final asset price.

This poor estimation of the mean indicates that if the time horizon T' of sample path for block bootstrap resampling is not sufficiently long, there exists a significant path dependent distance between the true and bootstrap resampled distribution of total log return. For that of the Customized AR(1) model, the resampled means show similar results.

b_{op}	Theoretical	Block size				
		1	b_{op}	5	10	20
1.7934	-0.3000	0.5670	0.5702	0.5784	0.5806	0.5792
2.3485	-0.3000	0.2645	0.2634	0.2581	0.2612	0.2626
3.4183	-0.3000	-0.1796	-0.1820	-0.1854	-0.1889	-0.1852
4.0068	-0.3000	1.0844	1.0872	1.0864	1.0936	1.0969

Table 5.1: Mean estimation for bootstrap resampled resampled total log return $R_{bootstrap,T}$ before shifting for different sample paths under the Geometric Brownian Motion, yielding $b_{op} = 1.7934, 2.3485, 3.4183, 4.0068$.

To mitigate the effect on mean but keep the shape of distribution consistent, we introduce a shift distance of return d_r, which is the optimal shift distance where the true and shifted distribution experience the largest similarity. Then, we shift the bootstrap resampled distribution towards the true distribution along the x-axis with distance d_r, and obtain the new bootstrap resampled distribution of return R_{shift}. After that, we will compare the true and shifted distribution.

First, we analyze the distribution of $R_{bootstrap,T}$ in the range $[r_{min}, r_{max}]$. We set N_{bin} as the number of bins, and define

$$\Delta r = (r_{max} - r_{min})/N_{bin},$$

$$r_1 = r_{min},\ r_{i+1} = r_i + \Delta r,$$

and the bin edge r as

$$r = \{r_i,\ i \in [1, N_{bin} + 1]\}.$$

We define the probability density function for one distribution, X, as

$$X = \{X_i,\ i \in [1, N_{bin} + 1]\},$$

where X_i is the probability of X falling in the i^{th} bin. Similarly, let Y be the probability of another distribution for the same definition of bins.

To calculate the distance between X and Y, we introduce the correlation function $C(d)$ as

$$C(d) = \begin{cases} \sum_{i=1}^{N-d} X_i Y_{i+d} & d \geqslant 0 \\ C(-d) & d \leq 0 \end{cases} \tag{5.3}$$

to measure the cross-correlation, which could also treated as the similarity, between X and shifted Y as a function of the shift d. The optimal shift $d^\star$ is defined as

$$d^\star = \arg\max_{\mathbf{d}} C(d), \tag{5.4}$$

which is the distance to achieve the largest overlapped area between the true and bootstrap resampled distribution. Figure 5.1 shows the plot of cross-correlation $C(d)$ for bootstrap resampled distribution based on a sample path of a Geometric Brownian Motion whose $b_{op} = 1.7934$, and for other situations the plots are in the same shape.

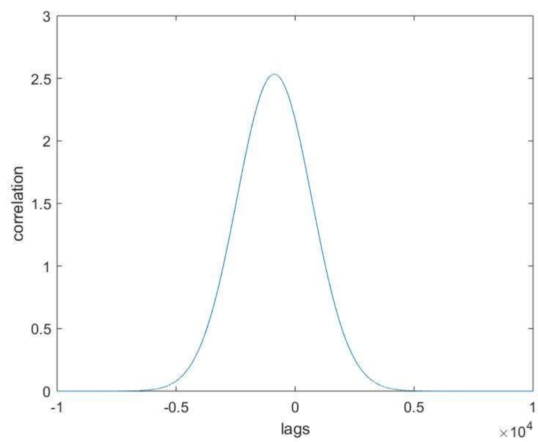

Figure 5.1: Plot of correlation $C(d)$ with lag d for a bootstrap resampled distribution of a sample path of a Geometric Brownian Motion whose $b_{op} = 1.7934$

In this case, the optimal lag $d^\star$ is the number of bins to be shifted. To eliminate its effect on $d^\star$, the number of bins N_{bin} in the range $[r_{min}, r_{max}]$ is chosen to be sufficiently large. In addition, the value should be proportional to the number of simulation N_{sim} to ensure that there is enough data to calculate the cross-correlation for each interval on average.

Finally, to obtain the shifting distance in terms of returns, we define

$$d_r = d^\star \times \Delta r,$$

$$R_{shift} = R_{bootstrap} + d_r,$$

and we get the bootstrap resampled distribution after shifting Y_{shift} simultaneously. It will be shown in the further computation that the resampled distributions before and after shifting share the distribution statistics such as variance, skewness and kurtosis that are related to the shape of distribution.

Chapter 6

Bootstrap resampling analysis

In this research, we conduct computation for total log return under the Geometric Brownian Motion and the Customized AR(1) model, and evaluate the performance of bootstrap resampled distributions in comparison to the true distribution. We set the target time horizon $T = 30$ years, time horizon for a sample observation path $T' = 100$ years, $\Delta t = 1/252$ year and the number of simulations $N_{sim} = 200,000$. We simulate the true distribution based on the discretized stochastic processes, as discribed in Algorithm 1 and 2 for each model. We generate multiple sample paths of the monthly log returns, yielding different optimal block sizes b_{op}, and choose expected block sizes for block bootstrap resampling following the procedure in Algorithm 3. When computing the distance of distribution of return, we choose the range $[r_{min}, r_{max}] = [-5, 5]$, $N_{bin} = 10,000$, and $\Delta r = 0.001$.

For the Geometric Brownian Motion, we focus on the performance of block bootstrap in the estimation of the distribution for monthly and total log return. For the Customized AR(1) model, in addition to the bootstrap resampled distribution, we also discuss the performance of bootstrap resampling estimation for variance ratio $VR(k)$.

6.1 Performance under the Geometric Brownian Motion

For the Geometric Brownian Motion, we use the inputs in Table 4.1. We choose 4 sample paths yielding an optimal block size $b_{op} = 1.7934$, 2.3485, 3.4183, 4.0668. For each sample path, we obtain a $200,000 \times 360$ matrix after block bootstrap resampling, where every resampling path represents a series of 30-year monthly log return.

6.1.1 Distributions of monthly log returns

In this section, we discuss the bootstrap resampled monthly return $R_{bootstrap,month}$ with the stationary block bootstrap resampling. Table 6.1 shows the mean and variance of bootstrap resampled monthly return from the four sample paths above. It is found that the difference of estimations using different block sizes is not significant. Compared with the theoretical value, the variance estimation is within the error tolerance of 5%. On the other hand, the accuracy in the mean estimator is poor. This is consistent with the test in the previous chapter. We need to further correct the mean estimator.

($b_{op} = 1.7934$)

Statistics	Theoretical	Block size				
		1	b_{op}	5	10	20
Mean	-0.0008	0.0016	0.0016	0.0016	0.0016	0.0016
Var	0.0033	0.0035	0.0035	0.0035	0.0035	0.0035

($b_{op} = 2.3485$)

Statistics	Theoretical	Block size				
		1	b_{op}	5	10	20
Mean	-0.0008	0.0007	0.0007	0.0007	0.0007	0.0007
Var	0.0033	0.0032	0.0032	0.0032	0.0032	0.0032

($b_{op} = 3.4183$)

Statistics	Theoretical	Block size				
		1	b_{op}	5	10	20
Mean	-0.0008	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005
Var	0.0033	0.0033	0.0033	0.0033	0.0033	0.0033

($b_{op} = 4.0068$)

Statistics	Theoretical	Block size				
		1	b_{op}	5	10	20
Mean	-0.0008	0.0030	0.0030	0.0030	0.0030	0.0030
Var	0.0033	0.0032	0.0032	0.0032	0.0032	0.0032

Table 6.1: Statistics for bootstrap resampled monthly log return $R_{bootstrap,month}$ for different sample paths under the Geometric Brownian Motion, yielding $b_{op} = 1.7934$, 2.3485, 3.4183, 4.0668.

6.1.2 Distributions of total log returns

Standard bootstrap resampling

Based on (3.7), the log return of the Geometric Brownian Motion follows a standard Brownian motion (SBM) which indicates that in the sample path R_{sample}, $cov(R_{sample,i}, R_{sample,j}) = 0, i, j \in [1, N]$, where N is the length of resampling paths. Since SBM is i.i.d, the optimal block size for block bootstrap resampling should be one. Thus, we first conduct block bootstrap resampling with a fixed block size of one month, which is the standard bootstrap resampling. We compare the distribution results for total log return $R_{bootstrap,T}$ for different sample paths.

Table 6.2 shows the statistics of block bootstrap resampled distributions before and after shifting for different sample paths. It is found that variance, skewness and kurtosis are consistent before and after shifting, while mean is corrected within the estimation error of 5% compared with the theoretical values. In addition, all statistics are sensitive to the sample path. These phenomena are also expected with stationary block bootstrap resampling.

Total return	Theoretical	Sample paths			
		1.7934	2.3485	3.4183	4.0668
Before shifting					
Mean	-0.3000	0.5702	0.2601	-0.1770	1.0856
Var	1.2000	1.2549	1.1516	1.1965	1.1592
Skewness	0.0000	-0.0057	-0.0053	-0.0009	-0.0050
Kurtosis	3.0000	2.9851	3.0063	3.0009	3.0158
d_r	-	-0.8740	-0.5630	-0.1260	-1.3880
After shifting					
Mean	-0.3000	-0.3038	-0.3029	-0.3030	-0.3024
Var	1.2000	1.2549	1.1516	1.1965	1.1592
Skewness	0.0000	-0.0057	-0.0053	-0.0009	-0.0050
Kurtosis	3.0000	2.9851	3.0063	3.0009	3.0158

Table 6.2: Statistics for bootstrap resampled total log return $R_{bootstrap,T}$ for different sample paths under the Geometric Brownian Motion, yielding $b_{op} = 1.7934$, 2.3485, 3.4183, 4.0668. The block size is fixed as one month.

Stationary block bootstrap resampling

In this section, we conduct stationary block bootstrap resampling and compare the difference between true and bootstrap resampled distributions under the Geometric Brownian Motion. We select the same sample paths in the previous section and the expected block sizes $b_{exp} = 1, b_{op}, 5, 10, 20$. We use the inputs displayed in Table 4.1. Figure 6.1 to 6.2 and Table 6.3 to 6.4 display the plots and statistics of bootstrap resampled distributions before and after shifting, with the comparison of theoretical values and these of the true distribution. From the plots and statistics, it is found that the bootstrap resampled results are block size dependent and path dependent. Since some statistics will be corrected after shifting, we compare the shifted distribution with the true distribution. We set the tolerance of estimation error as 5%.

Within the same sample path, the mean of R_{shift} is within an error of 4.22%, 2.21%, 3.31%, 3.29% compared with the true value of R_T. For the median, the errors are 2.77%, 2.17%, 2.61%, 2.17%. We conclude that the distribution after shifting performs well in the estimation of mean and median. For standard deviation, with the increase of the expected block size, the value decreases significantly. In general, skewness and kurtosis move in the opposite direction of the theoretical values. Since these three indicators remain consistent before and after shifting, the decreasing trend shows that the bootstrap resampled distribution changes away from a normal distribution when the block size increases. Meanwhile, it is found that the tails of bootstrap resampled distribution with larger block sizes are thinner than that with smaller block sizes. Also from the distribution plots, we find that the peak of bootstrap resampled distribution increases when block size increases.

We show that this decreasing trend is due to the property of block bootstrap resampling. We start with the block bootstrap resampling with fixed block size. We set the size of sample path as N, the block size as b, the number of blocks in each simulation path l, and the size of each resampling path n where $n = b \times l$. For block bootstrap resampling, each time we randomly choose a starting point in the sample path as the start of block, so the block bootstrap resampling could be treated as selecting l starting points from the sample path where the number of starting points is N. Then, since the same starting point could be selected repeatedly, the number of combinations for selecting starting points for one sample path is N^l. Since we know that the optimal block size for log return of Geometric Brownian Motion is one, so if block size $b_1 > b_2 > 1$, then $l_1 < l_2 < n$, $N^{l_1} < N^{l_2} < N^n$, hence using a block size larger than one will trigger the simulation paths of log return with higher similarity.

Since for stationary block bootstrap resampling, $E(b_{ac}) = b_{exp}$, we could link it with block bootstrap resampling with fixed block size, so this similar-sampling problem will also occur. As a result, when block size is larger, the sums of log return for different paths are more clustered around mean, hence the variance tends to be smaller, skewness and kurtosis diverges from the true value, and then the bootstrap resampled distribution becomes non-normal and the peak of distribution in the plots becomes steeper.

6.2 Performance under the Customized AR(1) model

In this section, we discuss the performance of the block bootstrap resampling in estimating the variance ratio and distribution of total log return under the Customized AR(1) model. In these two tests, we conduct the stationary block bootstrap resampling assuming different volatility ratios or different model parameters. We explore the trend of estimation with the variable changes. For the estimation of variance ratio, we also compare the estimation bias of different volatility ratios between continuous and discrete form in [6].

6.2.1 Estimation of variance ratio $VR(k)$

As showned in (3.24), the variance ratio $VR(k)$ is only related to the time horizon m, volatility ratio $r = \sigma_1/\sigma_2$, and model parameter θ_2. The estimation bias is defined as $E(\widehat{VR}(k)) - VR(k)$, where $\widehat{VR}(k)$ and $VR(k)$ are the corresponding estimated and theoretical value of variance ratio. In [6], the bootstrap resampled estimation bias of the

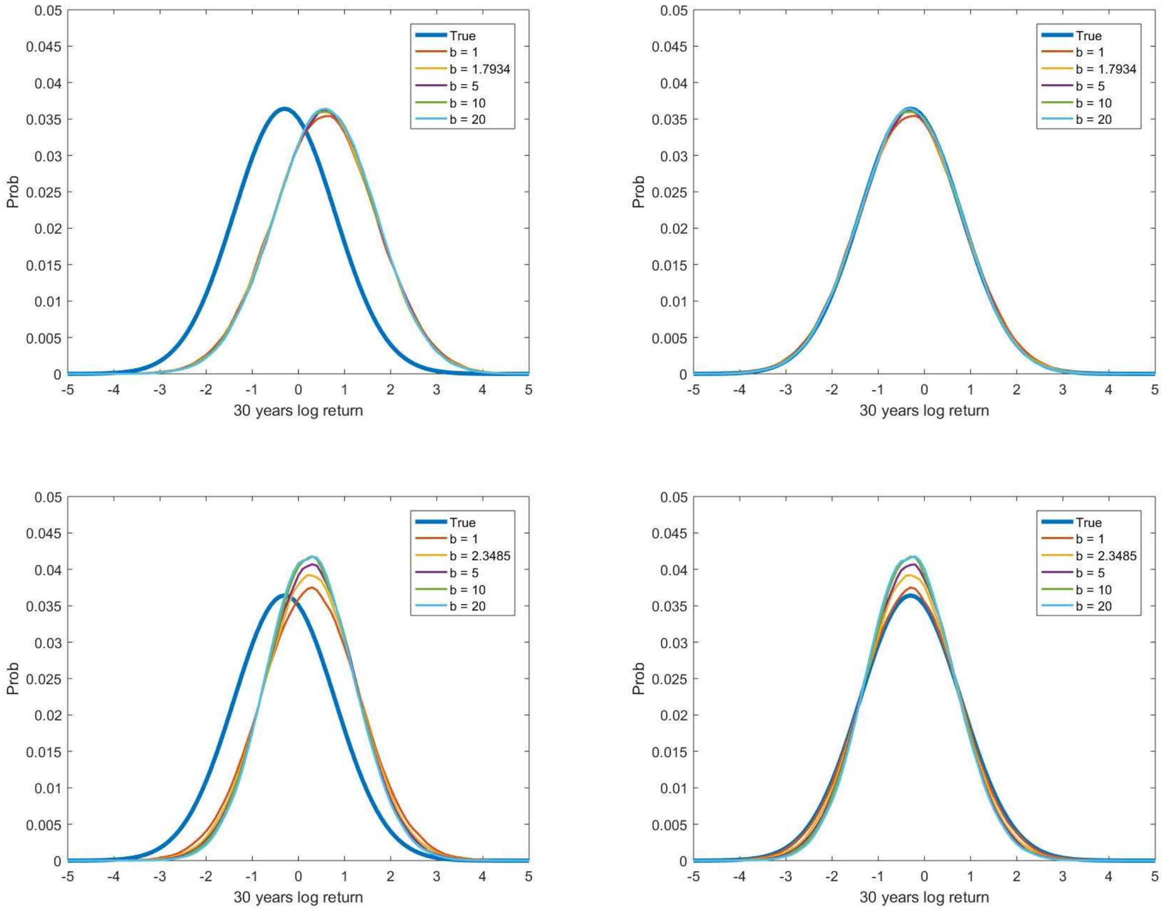

Figure 6.1: Bootstrap resampled distribution for total log return $R_{bootstrap,T}$ for different sample paths under the Geometric Brownian Motion, yielding $b_{op} = 1.7934$, 2.3485 (Left: before shifting; Right: after shifting).

variance ratio using the moving and stationary block bootstrap resampling was conducted. In this section, we conduct the same computation for the continuous form by stationary block bootstrap resampling. We compare the results with that in the previous research.

We simulate 5000 series of annual log returns and obtain the variance of single period log return $Var(X_i)$ for each simulation. After that we use block sizes b to implement stationary block bootstrap resampling for 1000 blocks, where the resampling series is a matrix with size $(1000b) \times 1$. We calculate the variance of k-period log return $Var(X_{i,i+k})$ as a moving sum and the estimation of variance ratio $VR(k)$. After the simulation, we obtain the mean of the bootstrap resampled $VR(k)$ for the specific time horizon and block size. In the computation, we select a series of k and b to determine the trend of estimation bias as the time horizon and block size change.

Following the same inputs of the test in the discrete form (3.9), we set model parameter $\theta_2 = 0.1$ ($\phi = 0.9$ in the discrete form), the volatility ratio $r \in \{0.0, 0.5, 1.0, 1.5\}$, and

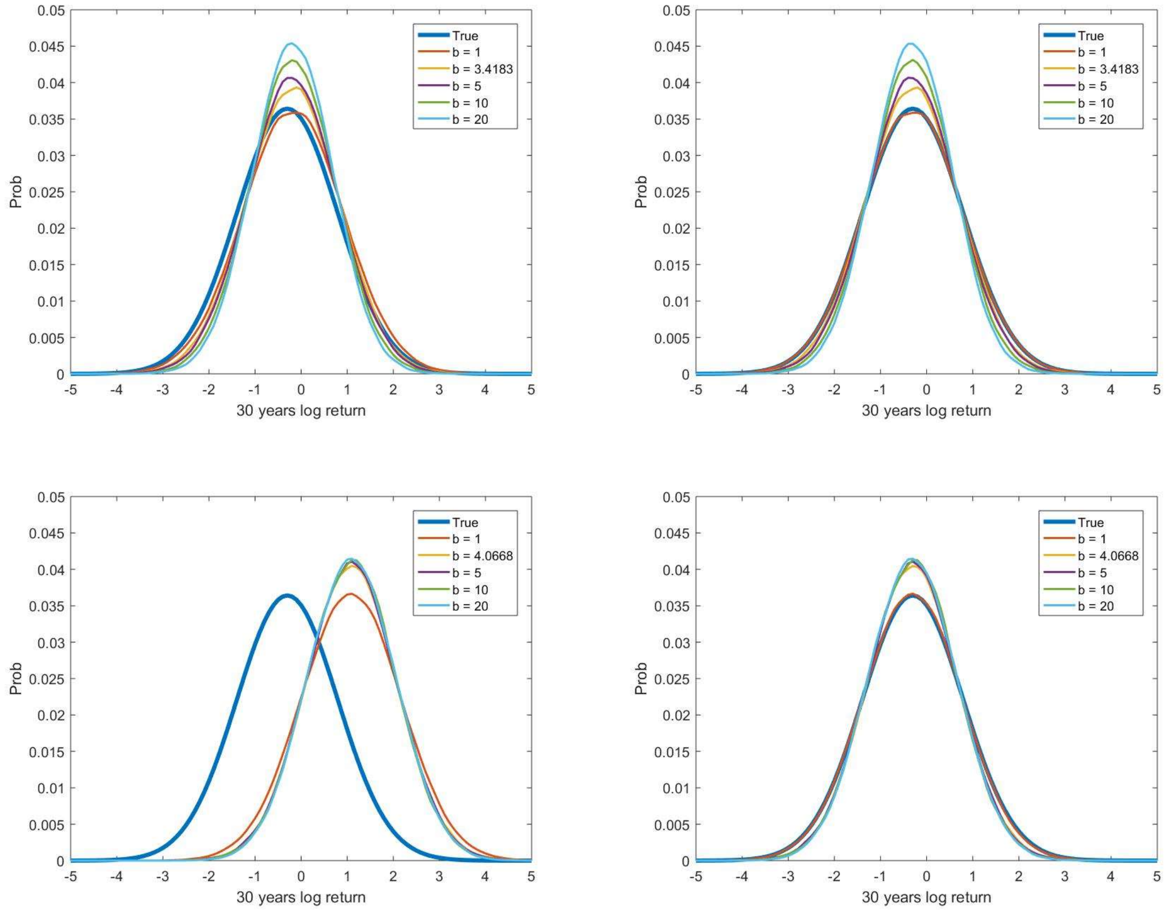

Figure 6.2: Bootstrap resampled distribution for total log return $R_{bootstrap,T}$ for different sample paths under the Geometric Brownian Motion, yielding $b_{op} = 3.4183$, 4.0068 (Left: before shifting; Right: after shifting).

the sample length $n \in \{63, 120\}$. We estimate the variance ratio with the time horizon $m \in \{2, 4, \ldots, 20\}$ years, and the expected block size used in stationary block bootstrap resampling $b_{exp} = k$ for each value of time horizon.

Table 6.5 shows the estimation bias of the variance ratio under the continuous model (3.18). It is found that, similar to the resampling result under the discrete model (3.9), these is an significant increasing trend of estimation bias in $VR(k)$ with increases of time horizon. This shows the existence of estimation error of block bootstrap resampling, which is consistent with that for the discrete form. This motivates the selection of the optimal block size to minimize the estimation bias in $VR(k)$ in [6]. Also, it is found that when σ_1/σ_2 tends to zero, the bias in the continuous form is slightly larger than that in the discrete form. When σ_1 dominates, the continuous form outperforms generally. According to [6], the reason to both positive and negative bias is due to the extra negative correlation induced by overlapped blocks (explained in [24, 26]), which is related to the volatility ratio.

($b_{op} = 1.7934$)

Statistics	Theoretical	Block size				
		1	b_{op}	5	10	20
		Before shifting				
Mean	-0.3000	0.5670	0.5702	0.5784	0.5806	0.5792
Var	1.2000	1.2544	1.2166	1.1890	1.2018	1.1774
Skewness	0.0000	-0.0030	-0.0035	-0.0197	-0.0183	-0.0069
Kurtosis	3.0000	2.9941	3.0108	2.9852	2.9828	2.9514
d_r	-	-0.8700	-0.8730	-0.8860	-0.8890	-0.8830
		After shifting				
Mean	-0.3000	-0.3030	-0.3028	-0.3076	-0.3084	-0.3038
Var	1.2000	1.2544	1.2166	1.1890	1.2018	1.1774
Skewness	0.0000	-0.0030	-0.0035	-0.0197	-0.0183	-0.0069
Kurtosis	3.0000	2.9941	3.0108	2.9852	2.9828	2.9514

($b_{op} = 2.3485$)

Statistics	Theoretical	Block size				
		1	b_{op}	5	10	20
		Before shifting				
Mean	-0.3000	0.2645	0.2634	0.2581	0.2612	0.2626
Var	1.2000	1.1439	1.0300	0.9512	0.9056	0.8795
Skewness	0.0000	-0.0097	0.0075	0.0074	0.0201	0.0436
Kurtosis	3.0000	3.0039	2.9893	2.9896	2.9819	2.9724
d_r	-	-0.5690	-0.5630	-0.5600	-0.5590	-0.5550
		After shifting				
Mean	-0.3000	-0.3045	-0.2996	-0.3019	-0.2978	-0.2924
Var	1.2000	1.1439	1.0300	0.9512	0.9056	0.8795
Skewness	0.0000	-0.0097	0.0075	0.0074	0.0201	0.0436
Kurtosis	3.0000	3.0039	2.9893	2.9896	2.9819	2.9724

Table 6.3: Statistics for bootstrapped resampled total log return $R_{bootstrap,T}$ for different sample paths under the Geometric Brownian Motion, yielding $b_{op} = 1.7934, 2.3485$.

The positive bias is due to the random block size generated by a geometric distribution during the stationary block bootstrap resampling, and this randomness destroys the serial dependence in the original data. Thus, the bias is the mitigation of two kinds of errors which might be positive or negative.

6.2.2 Distribution of monthly log return

In this section, we focus on the block bootstrap resampled monthly return $R_{bootstrap,month}$ with the stationary block bootstrap resampling. We use the inputs in Table 4.2 set ratio $r = 0, 1.5,\ \theta_2 = 0.1$ and $r = 1,\ \theta_2 = 0.1, 0.9$. Table 6.6 shows the block bootstrap resampled

($b_{op} = 3.4183$)

Statistics	Theoretical	Block size				
		1	b_{op}	5	10	20
Before shifting						
Mean	-0.3000	-0.1796	-0.1820	-0.1854	-0.1889	-0.1852
Var	1.2000	1.1945	1.0157	0.9611	0.8495	0.7555
Skewness	0.0000	-0.0056	-0.0068	0.0026	-0.0055	-0.0208
Kurtosis	3.0000	3.0161	3.0028	3.0134	3.0055	2.9946
d_r	-	-0.1230	-0.1210	-0.1160	-0.1140	-0.1200
After shifting						
Mean	-0.3000	-0.3026	-0.3030	-0.3014	-0.3029	-0.3052
Var	1.2000	1.1945	1.0157	0.9611	0.8495	0.7555
Skewness	0.0000	-0.0056	-0.0068	0.0026	-0.0055	-0.0208
Kurtosis	3.0000	3.0161	3.0028	3.0134	3.0055	2.9946

($b_{op} = 4.0068$)

Statistics	Theoretical	Block size				
		1	b_{op}	5	10	20
Before shifting						
Mean	-0.3000	1.0844	1.0872	1.0864	1.0936	1.0969
Var	1.2000	1.1532	0.9618	0.9507	0.9166	0.9070
Skewness	0.0000	-0.0150	-0.0020	0.0071	-0.0093	0.0120
Kurtosis	3.0000	2.9913	2.9913	3.0027	2.9914	2.9297
d_r	-	-1.3900	-1.3900	-1.3870	-1.3970	-1.3960
After shifting						
Mean	-0.3000	-0.3056	-0.3028	-0.3006	-0.3034	-0.2991
Var	1.2000	1.1532	0.9618	0.9507	0.9166	0.9070
Skewness	0.0000	-0.0150	-0.0020	0.0071	-0.0093	0.0120
Kurtosis	3.0000	2.9913	2.9913	3.0027	2.9914	2.9297

Table 6.4: Statistics for bootstrapped resampled total log return $R_{bootstrap,T}$ for different sample paths under the Geometric Brownian Motion, yielding $b_{op} = 3.4183, 4.0068$.

statistics of monthly return for these settings above. It is found that the block bootstrap resampling performs well in estimating the distribution statistics for monthly return, and the estimations for distribution statistics among different block sizes are the same within four decimal places.

n	ratio	Time horizon (years)									
		2	4	6	8	10	12	14	16	18	20
63	0.0	0.01	0.03	0.04	0.06	0.07	0.08	0.08	0.09	0.10	0.10
	0.5	0.00	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.04
	1.0	-0.01	0.00	-0.01	-0.01	-0.01	-0.01	-0.02	-0.03	-0.04	-0.05
	1.5	-0.01	-0.03	-0.03	-0.05	-0.06	-0.07	-0.09	-0.10	-0.11	-0.13
120	0.0	0.02	0.05	0.07	0.08	0.10	0.11	0.12	0.13	0.13	0.13
	0.5	0.01	0.04	0.05	0.06	0.07	0.07	0.08	0.08	0.09	0.09
	1.0	0.01	0.02	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.01
	1.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.02	-0.02

Table 6.5: Estimation bias for the variance ratio, where Bias $= E(\widehat{VR}(k)) - VR(k)$, where $VR(k)$ is the theoretical value.

6.2.3 Distribution of total log return

Different volatility ratios

In this section we focus on the performance of the stationary block bootstrap resampling in the distribution estimation of the total log return with different volatility ratios r. We follow the same inputs in Table 4.2 and set ratio $r = 0.0, 0.5, 1.0, 1.5$, $\theta_2 = 0.1$. Following the same procedure for GBM, we simulate one sample path for each parameter with expected block size $b_{exp} = \{1, 10, 20, 50, 100, 150\}$ and compare the difference in distribution statistics of different block sizes.

Table 6.7 and Figure 6.3 show the bootstrap resampled distribution with different r. It is found that, when r tends to zero, the peak of true distribution for R_T becomes higher and the performance of bootstrap resampled distribution is poorer. The difference from plots and variance in the table becomes significant. However, when the ratio r increases and tends to 1.5, although the statistics still differ significantly, we could observe from the plot that the bootstrap resampled distribution tends to the true distribution in shape.

We explain this phenomena from the auto-correlation of the log return with the change of the volatility ratio. When the model parameter θ_2 is fixed, based on Figure 3.2, as the volatility ratio decreases and even tends to zero, the white noise in the temporary component V_{t_n} dominates. The absolute value of the first order auto-correlation $AC1(k)$ increases with the increase of the time horizon k. It is significant compared with that of other ratios. To maintain the large correlation, it is required to use a large block size for the stationary block bootstrap resampling. A small block size will destroy this dependence, which triggers a large variance and the bootstrap resampled distribution in plots may not be treated as an estimation of the true distribution. When r increases and the white noise in the incremental component D_{t_n} dominates the variance, the auto-correlation recedes. From Figure 6.3, it is found that the bootstrap resampled performance is better than before. Thus, it is expected that when r tends to infinity, the log return follows a Standard Brownian Motion and the optimal block size tends to one.

($r = 0.0,\ \theta_2 = 0.1$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
Mean	0.0008	0.0007	0.0008	0.0008	0.0008	0.0008	0.0008
Var	0.0033	0.0036	0.0036	0.0036	0.0036	0.0036	0.0036

($r = 1.5,\ \theta_2 = 0.1$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
Mean	0.0008	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
Var	0.0108	0.0107	0.0107	0.0107	0.0107	0.0107	0.0107

($r = 1.0,\ \theta_2 = 0.1$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
Mean	0.0008	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
Var	0.0108	0.0107	0.0107	0.0107	0.0107	0.0107	0.0107

($r = 1.0,\ \theta_2 = 0.9$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
Mean	0.0008	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
Var	0.0108	0.0107	0.0107	0.0107	0.0107	0.0107	0.0107

Table 6.6: Statistics for bootstrap resampled monthly log return $R_{bootstrap,month}$l for different sample paths under the Customized AR(1) mode, where $r = 0, 1.5, \theta_2 = 0.1$ and $r = 1, \theta_2 = 0.1, 0.9$.

Different model parameters

In this section we focus on the performance of the stationary block bootstrap resampling in the distribution estimation of the total log return with different volatility ratios r. We follow the same inputs in Table 4.2 and set $\theta_2 = 0.0, 0.1, 0.3, 0.5, 0.7, 0.9,\ r = 1$. Following the same procedure for GBM, we simulate one sample path for each parameter with expected block size $b_{exp} = \{1, 10, 20, 50, 100, 150\}$ and compare the difference in distribution statistics of different block sizes.

Figure 6.4 to 6.5 and Table 6.8 to 6.9 show the plots and statistics of the bootstrap resampled distributions for different values of θ_2. It is found that with $\theta_2 = 0$, the bootstrap resampled estimation is the similar to the Geometric Brownian Motion and the optimal block size is one month. However, when $\theta_2 > 0$ the estimation bias is much larger than tolerance 5% which indicates the bootstrap resampled distribution is poor in approximating the trend of the true distribuion. However, we still find the decreasing trend of the variance as the block size increases. From Table 6.9 it is found that for those two sample paths when $\theta_2 = 0.7$ and $\theta_2 = 0.9$, the optimal block sizes for variance estimation are within the range (50, 100) months. Outside this range, the variance diverges.

We believe that the phenomena are due to the dependence of the log return on with the change of model parameter. When $\theta_2 = 0$, based on (3.7) R_T becomes a Standard Brownian Motion with finite variance. The bootstrap resampled result is the same as that of GBM. However, when $\theta_2 > 0$, according to Figure 3.2, there exists correlation between the log returns. We find that the estimation of the bootstrap resampled distribution requires a larger block size to satisfy the estimation of true distribution. It is found that, in the long run, the absolute value of correlation converges to zero as θ_2 increases. Therefore, it is expected that the optimal block size decreases when θ_2 tends to zero and one.

6.3 Optimal block size

Based on the computation results for these two models, it is found that the block size influences the bootstrap resampled distribution of the total log return, no matter how large the serial correlation between sample set is. According to [1], the tradeoff of selecting a block size is described as follows. A small block size destroys the serial dependence between log returns. When the block size is too large, the bootstrap resampled paths become too similar. These two situations trigger a too-large and too-small variance, which verifies the trend of divergence for $\theta_2 = 0.7$ and $\theta_2 = 0.9$ in Table 6.9. More importantly, these results remind us the importance of selecting an optimal block size for block bootstrap resampling where serial correlation exists. This is also emphasized in [18, 14].

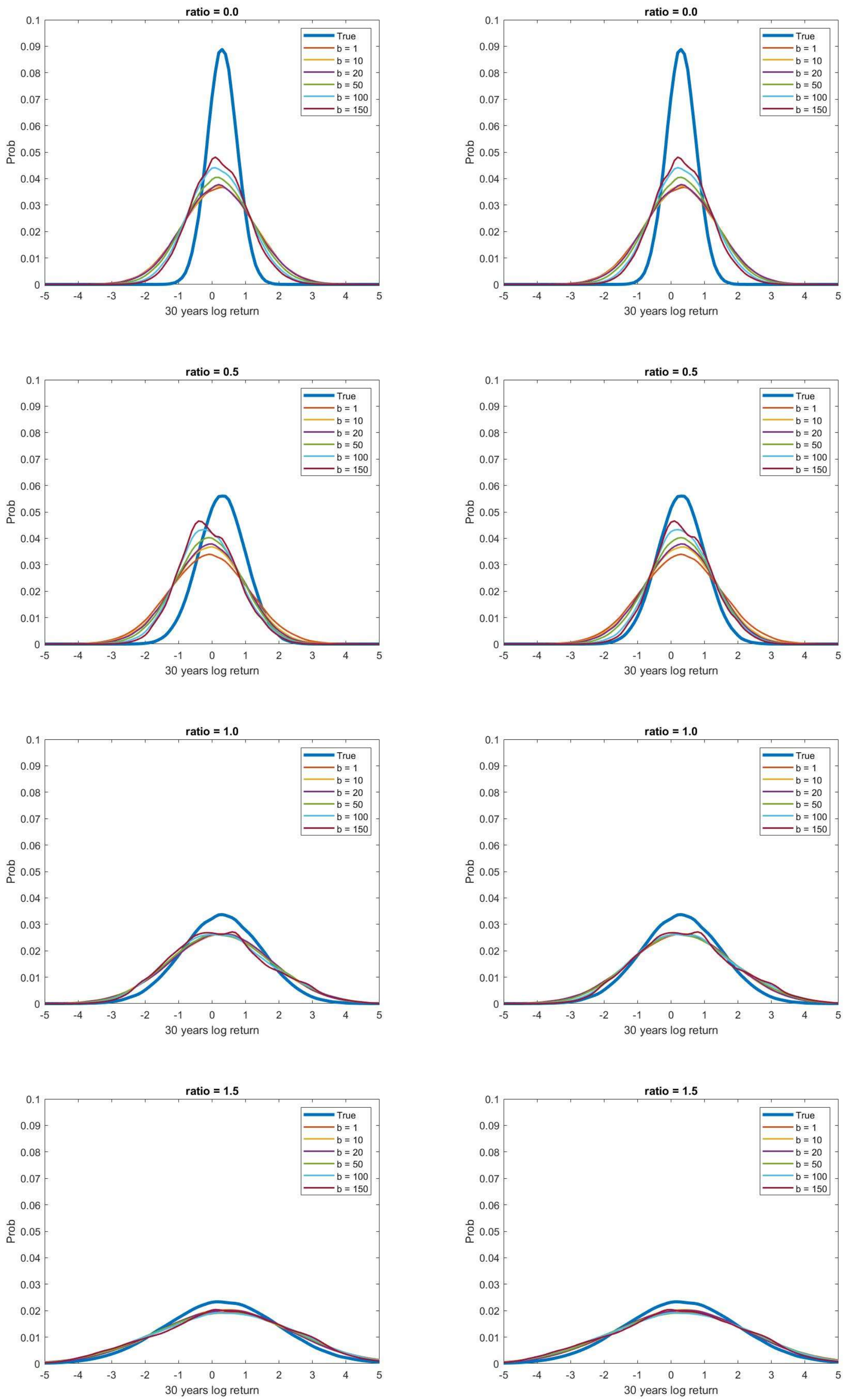

Figure 6.3: Bootstrap resampled distribution for total log return $R_{bootstrap,T}$ for different sample paths under the Customized AR(1) model where volatility ratio $r = 0.0, 0.5, 1.0, 1.5$ (Left: before shifting; Right: after shifting).

($r = 0.0$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
		Before shifting					
Mean	0.3009	0.1995	0.1996	0.2003	0.2011	0.1999	0.2019
Variance	0.1995	1.1846	1.1368	1.1328	0.9736	0.7962	0.7012
d_r	-	0.0930	0.1020	0.1160	0.1270	0.1370	0.1240
		After shifting					
Mean	0.3009	0.2925	0.3016	0.3163	0.3281	0.3369	0.3259
Variance	0.1995	1.1846	1.1368	1.1328	0.9736	0.7962	0.7012

($r = 0.5$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
		Before shifting					
Mean	0.3012	-0.0804	-0.0960	-0.1025	-0.0965	-0.1042	-0.1007
Variance	0.4995	1.3853	1.1719	1.0935	0.9605	0.8261	0.7417
d_r	-	0.3790	0.3710	0.3700	0.3870	0.4430	0.4540
		After shifting					
Mean	0.3012	0.2986	0.2750	0.2675	0.2905	0.3388	0.3533
Variance	0.4995	1.3853	1.1719	1.0935	0.9605	0.8261	0.7417

($r = 1.0$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
		Before shifting					
Mean	0.3000	0.2692	0.2712	0.2781	0.2804	0.2748	0.2708
Variance	1.3995	2.2651	2.2248	2.2169	2.2176	2.1707	2.0969
d_r	-	0.0260	0.0230	0.0210	0.0790	0.1540	0.1770
		After shifting					
Mean	0.3000	0.2952	0.2942	0.2991	0.3594	0.4288	0.4478
Variance	1.3995	2.2651	2.2248	2.2169	2.2176	2.1707	2.0969

($r = 1.5$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
		Before shifting					
Mean	0.3009	0.3711	0.3803	0.3784	0.3715	0.3661	0.3720
Variance	2.8995	4.0756	3.7906	3.7875	4.0612	4.0275	3.8350
d_r	-	0.0670	0.0880	0.0870	0.0890	0.1250	0.1580
		After shifting					
Mean	0.3009	0.3041	0.2923	0.2914	0.2825	0.2411	0.2140
Variance	2.8995	4.0756	3.7906	3.7875	4.0612	4.0275	3.8350

Table 6.7: Statistics for bootstrap resampled total log return $R_{bootstrap,T}$ for different sample paths under the Customized AR(1) model, where $r = 0.0, 0.5, 1.0, 1.5$.

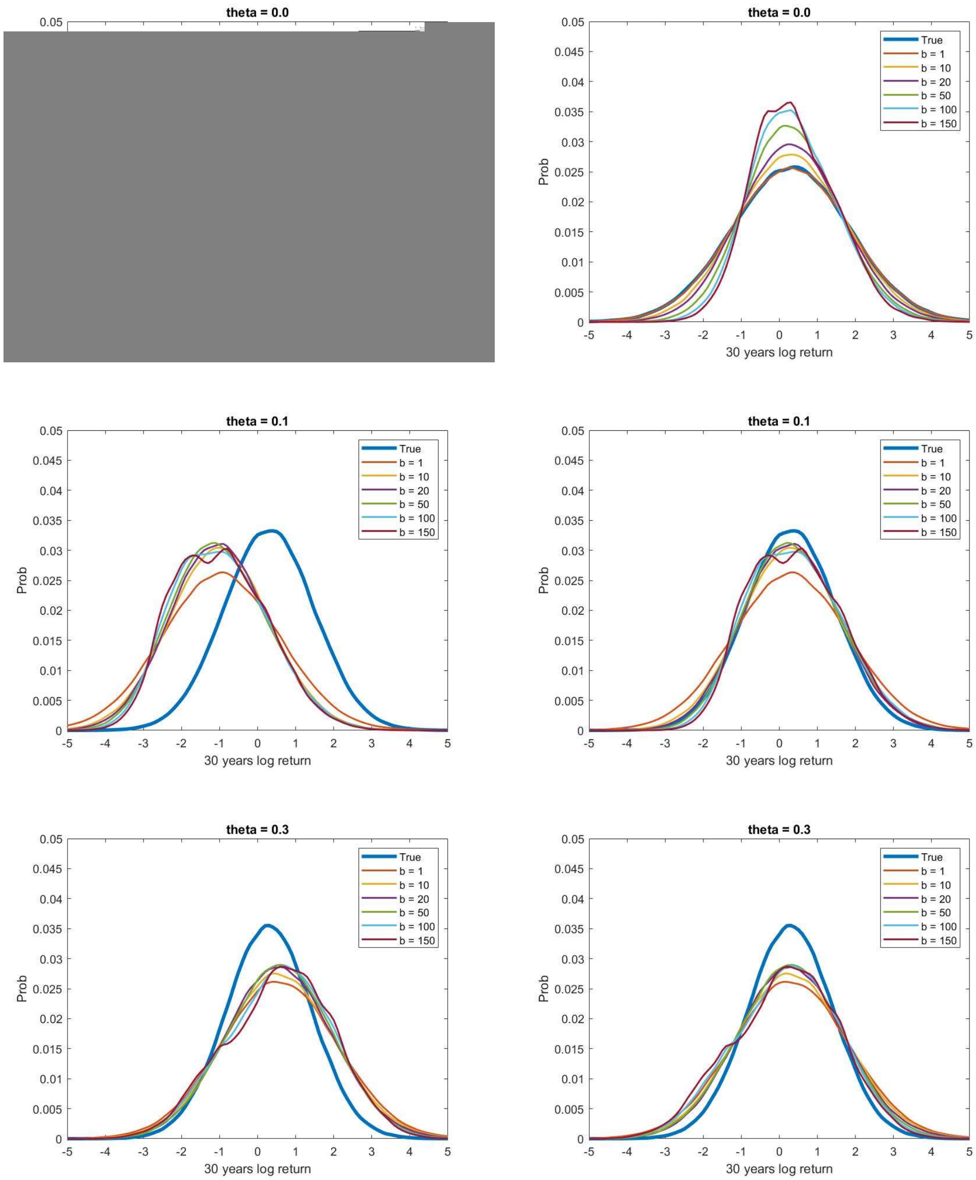

Figure 6.4: Bootstrap resampled distribution for total log return $R_{bootstrap,T}$ for different sample paths under the Customized AR(1) model where $\theta_2 = 0.0, 0.1, 0.3$ (Left: before shifting; Right: after shifting).

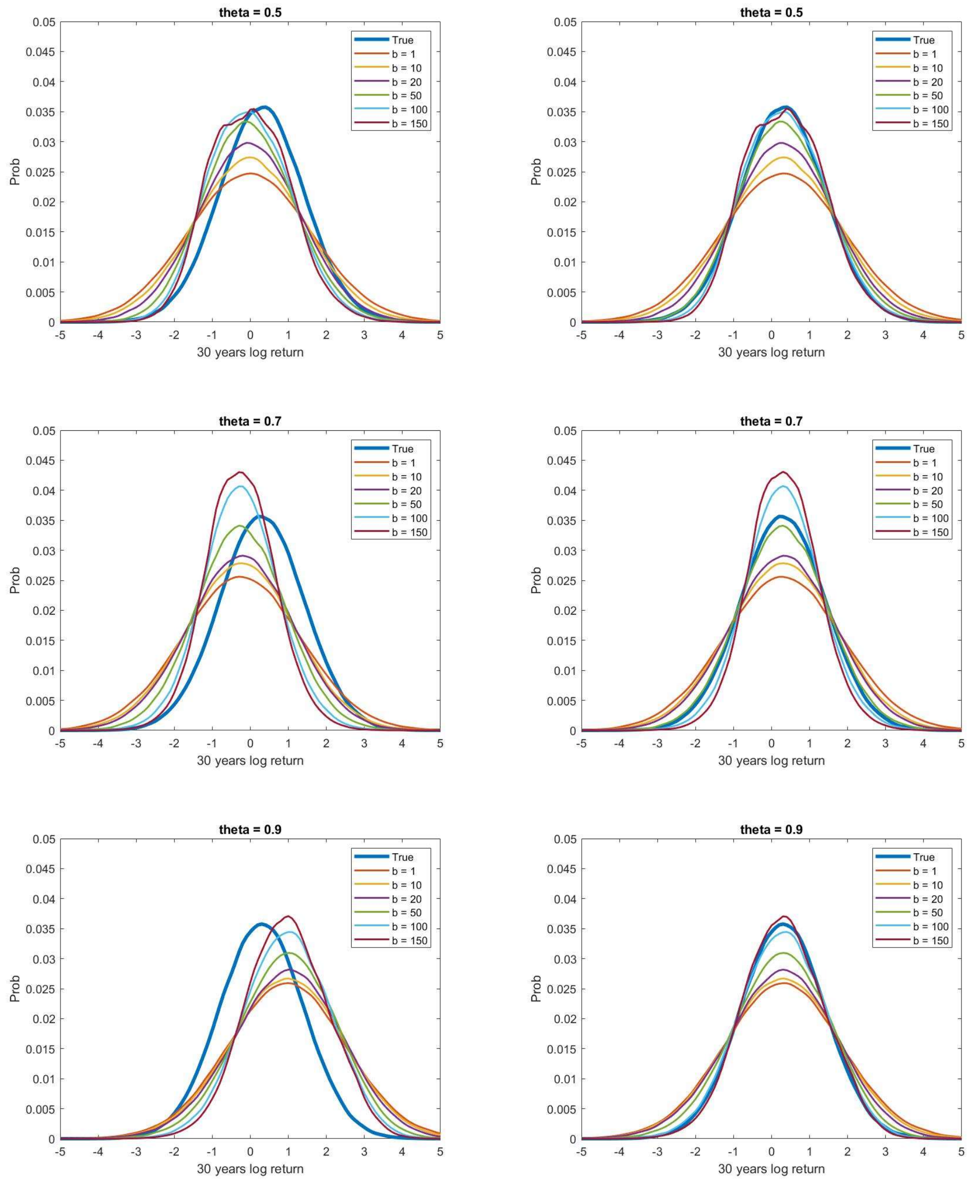

Figure 6.5: Bootstrap resampled distribution for total log return $R_{bootstrap,T}$ for different sample paths under the Customized AR(1) model where $\theta_2 = 0.5, 0.7, 0.9$ (Left: before shifting; Right: after shifting).

($\theta_2 = 0.0$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
Before shifting							
Mean	0.3000	0.4044	0.4256	0.4174	0.4187	0.4195	0.4181
Variance	2.4000	2.4030	2.0240	1.8091	1.4848	1.2824	1.1835
d_r	-	-0.0990	-0.1020	-0.0890	-0.0710	-0.0560	-0.0470
After shifting							
Mean	0.3000	0.3054	0.3236	0.3284	0.3477	0.3635	0.3711
Variance	2.4000	2.4030	2.0240	1.8091	1.4848	1.2824	1.1835

($\theta_2 = 0.1$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
Before shifting							
Mean	0.3000	-0.9601	-0.9912	-0.9984	-0.9965	-0.9962	-0.9962
Variance	1.3995	2.3190	1.7094	1.6091	1.5782	1.5343	1.4774
d_r	-	1.2670	1.3050	1.3300	1.3770	1.4010	1.4060
After shifting							
Mean	0.3000	0.3069	0.3138	0.3316	0.3805	0.4048	0.4098
Variance	1.3995	2.3190	1.7094	1.6091	1.5782	1.5343	1.4774

($\theta_2 = 0.3$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
Before shifting							
Mean	0.3000	0.5669	0.5644	0.5723	0.5728	0.5706	0.5688
Variance	1.2667	2.2992	2.0846	1.9126	1.8418	1.9114	1.9362
d_r	-	-0.2700	-0.2540	-0.2590	-0.2940	-0.3650	-0.4170
After shifting							
Mean	0.3000	0.2969	0.3104	0.3133	0.2788	0.2056	0.1518
Variance	1.2667	2.2992	2.0846	1.9126	1.8418	1.9114	1.9362

Table 6.8: Statistics for bootstrapped resampled total log return $R_{bootstrap,T}$ for different sample paths under the Customized AR(1) model where $\theta_2 = 0.0, 0.1, 0.3$.

($\theta_2 = 0.5$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
		Before shifting					
Mean	0.3000	-0.0195	-0.0274	-0.0198	-0.0193	-0.0135	-0.0121
Variance	1.2400	2.5660	2.1524	1.7821	1.4186	1.1878	1.0890
d_r	-	0.3230	0.3260	0.3250	0.3620	0.3650	0.3570
		After shifting					
Mean	0.3000	0.3035	0.2986	0.3052	0.3427	0.3515	0.3449
Variance	1.2400	2.5660	2.1524	1.7821	1.4186	1.1878	1.0890

($\theta_2 = 0.7$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
		Before shifting					
Mean	0.3000	-0.2348	-0.2615	-0.2580	-0.2472	-0.2543	-0.2539
Variance	1.2286	2.4183	2.0321	1.8537	1.3885	1.0008	0.8257
d_r	-	0.5380	0.5460	0.5430	0.5540	0.5670	0.5700
		After shifting					
Mean	0.3000	0.3032	0.2845	0.2850	0.3068	0.3127	0.3161
Variance	1.2286	2.4183	2.0321	1.8537	1.3885	1.0008	0.8257

($\theta_2 = 0.9$)

Statistics	Theoretical	Block size					
		1	10	20	50	100	150
		Before shifting					
Mean	0.3000	0.9578	0.9654	0.9668	0.9617	0.9601	0.9591
Variance	1.2222	2.3725	2.2006	2.0128	1.6342	1.3078	1.1513
d_r	-	-0.6610	-0.6840	-0.7140	-0.7020	-0.6700	-0.6460
		After shifting					
Mean	0.3000	0.2968	0.2814	0.2528	0.2597	0.2901	0.3131
Variance	1.2222	2.3725	2.2006	2.0128	1.6342	1.3078	1.1513

Table 6.9: Statistics for bootstrap resampled total log return $R_{bootstrap,T}$ for different sample paths under the Customized AR(1) model where $\theta_2 = 0.5, 0.7, 0.9$.

Chapter 7

Conclusions

In this research, we evaluate the performance of the block bootstrap resampling, especially stationary block bootstrap resampling in the estimation of total log return for a given time horizon. We use two models for the asset price, the Geometric Brownian Motion and the Customized AR(1) model. We derive the continuous form for the Customized AR(1) model, and calculate distribution statistics of the log return based on the continuous form of these two models. After simulating a sample path of monthly return, we implement the stationary block bootstrap resampling to obtain the total log return for the target time horizon.

We find that for the Geometric Brownian Motion, since its log return follows a random walk, the optimal block size is one. The increase of the block size will trigger high similarity of simulation paths and incur a small bootstrap variance. For the Customized AR(1) model, significant auto-correlation explains that the small block size could not satisfy the estimation requirement to approximate the true distribution of total log return. Based on the definition in [20], we conclude that the block bootstrap resampling performs well in approximating level-1 parameters, however for level-2 and high-level parameters, it is important to choose an optimal block size for estimating the statistics of interest.

There are some directions for future works. One potential aspect is to find the optimal block size directly based on the purpose of minimizing the variance or other distribution statistics which could represent the bootstrapped distribution of the total log return. Another aspect is that we could calculate the distance of distributions such as the Wasserstein distance introduce in [32], which is the minimal cost to transfer the bootstrapped distribution to the true distribution. Both these two methods are non plug-in data-driven methods. What is more, since we correct the bootstrapped distribution with the shifting distance, it is possible to test the performance of block bootstrap resampling return in further analysis such as asset allocation and risk forecasting.

References

[1] J. Berkowitz and L. Kilian. Recent developments in bootstrapping time series. *Econometric Reviews*, 19(1):1–48, 2000.

[2] P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217, 1981.

[3] P. Bühlmann and H. R. Künsch. The blockwise bootstrap for general parameters of a stationary time series. *Scandinavian Journal of Statistics*, 22(1):35–54, 1995.

[4] E. Carlstein. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 14(3):1171–1179, 1986.

[5] P. Cogneau and V. Zakamouline. Bootstrap methods for finance: Review and analysis. Working paper, University of Liege, 2010.

[6] P. Cogneau and V. Zakamouline. Block bootstrap methods and the choice of stocks for the long run. *Quantitative Finance*, 13(9):1443–1457, 2013.

[7] H. Dichtl, W. Drobetz, and M. Wambach. Testing rebalancing strategies for stock-bond portfolios across different asset allocations. *Applied Economics*, 48(9):772–788, 2016.

[8] H. Dichtl, W. Drobetz, and M. Wambach. A bootstrap-based comparison of portfolio insurance strategies. *The European Journal of Finance*, 23(1):31–59, 2017.

[9] B. Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.

[10] E. F. Fama and K. R. French. Permanent and temporary components of stock prices. *Journal of Political Economy*, 96(2):246–273, 1988.

[11] J. Fleming, C. Kirby, and B. Ostdiek. Bootstrap tests of multiple inequality restrictions on variance ratios. *Economics Letters*, 91(3):343–348, 2006.

[12] P. Forsyth and K. R. Vetzal. Optimal asset allocation for retirement saving: deterministic vs. time consistent adaptive strategies. *Applied Mathematical Finance*, 26(1):1–37, 2019.

[13] C. M. Garcia, P. R. Jackson, and M. H. Garcia. Confidence intervals in the determination of turbulence parameters. *Experiments in Fluids*, 40(4):514–522, 2006.

[14] P. Hall, J. L. Horowitz, and B. Y. Jing. On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574, 1995.

[15] K. He. The robustness of bootstrap estimator of variance. *Journal of the Italian Statistical Society*, 4(2):183–193, 1995.

[16] B. N. Huang. Do Asian stock market prices follow random walks? Evidence from the variance ratio test. *The Review of Financial Studies*, 1(1):41–66, 1988.

[17] J. H. Kim. Automatic variance ratio test under conditional heteroskedasticity. *Finance Research Letters*, 6(3):179–185, 2009.

[18] H. R. Kunsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1261, 1989.

[19] S. N. Lahiri. Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404, 1999.

[20] S. N. Lahiri. *Resampling methods for dependent data*. Springer, New York, 2003.

[21] S. N. Lahiri, K. Furukawa, and Y. D. Lee. A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. *Statistical Methodology*, 4(1=3):292–321, 2007.

[22] R. Y. Liu and K. Singh. Moving blocks jackknife and bootstrap capture weak dependence. *In: R. Lepage and L. Billard, Eds., Exploring the Limits of Bootstrap*, 1992.

[23] A. Lo and A. Mackinlay. Stock market prices do not follow random walks: evidence from a simple specification test. *The Review of Financial Studies*, 1(1):41–66, 1988.

[24] F. H. C. Marriott and J. A. Pope. Bias in the estimation of autocorrelations. *Biometrika*, 41(3/4):390–402, 1954.

[25] J. F. Muth. Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, 55(290):299–306, 1960.

[26] G. H. Orcutt and J. O. Irwin. A study of the autoregressive nature of the time series used for Tinbergen's model of the economic system of the United States, 1919-1932. *Journal of the American Statistical Association*, 10(1):1–45, 1948.

[27] D. N. Politis, A. Patthon, and H. White. Correction to 'Automatic block-length selection for the dependent bootstrap' by D. Politis and H. White. *Econometric Reviews*, 28(4):372–375, 2009.

[28] D. N. Politis and J. P. Romano. The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, 1994.

[29] D. N. Politis and H. White. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

[30] K. Reddy and V. Clinton. Simulating stock prices using Geometric Brownian Motion: evidence from Australian companies. *Australasian Accounting, Business and Finance Journal*, 10(3):23–47, 2016.

[31] K. Singh. On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, 9(6):1187–1195, 1981.

[32] C. Villani. *Optimal Transport: Old and New*. Springer, Berlin Heidelberg, 2009.