AGENT-BASED MODELLING FOR FINANCIAL MARKETS WITH HETEROGENEOUS AGENTS: CALIBRATION AND SIMULATION

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Science and Engineering

2022

Yu Wang
Department of Mathematics
School of Natural Sciences

Contents

Word count 34478

List of Tables

List of Figures

The University of Manchester

Yu Wang

Doctor of Philosophy

Agent-based modelling for financial markets with heterogeneous agents: calibration and simulation

May 13, 2022

This thesis introduces a framework of using agent-based models (ABM) with heterogeneous agents to produce an artificial financial market. The first ABM develops an order drive market embedded with an endogenous imitation mechanism and a population of noisy agents. The model network structure is established that each noisy agent can be imitated by others with a probability proportional to his wealth. We investigate the influence of imitation behaviour on the statistical properties of model outputs, such as log-return and bid-ask spread. The presence of stylised empirical facts, like volatility clustering and zero auto-correlation in price return, are verified. A variant model is presented by replacing the noisy agents with trending-following ones who can use the Bayesian learning method to track the asset return. Agents place limit orders based on the optimal strategies derived from a portfolio choice problem. A simulated price crash can be caused by a positive feedback loop arising from learning behaviours. Meanwhile, a parameter sensitivity analysis is implemented for both ABMs, which helps choose the initial condition and a further model calibration. We also provide a framework for calibrating an ABM using a genetic algorithm, where we aim to find a region close to an optimal value due to the complexity and stochasticity of an ABM.

Declaration

No portion of the work referred to in the dissertation has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

Foremost, I would like to express my sincere to my supervisors, Dr Paul Johnson and Dr Saralees Nadarajah, for their comprehensive guidance, valuable advice, warm encouragement and continuous support throughout my PhD study. Their wealth of knowledge, creative ideas and critical thinking inspired me to be an independent researcher. I am extremely thankful for their invaluable assistance and I could not imagine completing this thesis without them.

I would also like to thank other members of the Mathematical Department, including Dr Kees van Schaik and Dr Sebastian Herrmann, who offered me precious comments and advice on my thesis writing and the project.

This thesis would not have been an enjoyable experience if I did not share an office with a fantastic group of friends. Many thanks to Song Xiao, Zubier, Lubomir, Karolis, Martyn, Jakey and Rajenki for the good time being together and the priceless relationship between us.

Throughout my four years of life in Manchester, I have made some great friends both in and out of the department. They have all made my time here in Manchester an enjoyable one.

I would also like to give my deepest thanks to my family. Without their unlimited support and love, I would have never accomplished this journey. To my parents, it seems only right that I dedicate this thesis to you

Last but not least, I would like to acknowledge the University of Manchester for its support throughout my PhD, especially in the period the world are fighting for the pandemic.

Chapter 1

Introduction

Understanding and modelling the behaviour of prices in financial markets have been attracting the attention of researchers for as long as financial markets have existed. Many empirical studies have made great efforts to explore and explain the dynamics of components in financial markets. Sharpe (1964) and Lintner (1965) introduced the capital asset pricing model, which states the decision about what rate of return of an asset is required theoretically to add to a well-diversified portfolio (see [1] and [2]). While Merton (1973) used several stochastic processes, such as jump processes and jump-diffusion processes, to model asset returns (see [3]). Later, Heath *et al.* (1992) proposed a bond pricing framework using a stochastic term structure of interest [4]. These papers attempt to deduce the price of major components, such as equities, interest rates and commodities given a market by modelling their dynamics with differential equations or stochastic differential equations. Meanwhile, the real-world financial markets are complex systems, where prices are usually generated from the actions and interactions of entities in the system, which has sparked a growing literature attempting to model the financial markets from the agent perspective. Motivated by this, the primary goal of this thesis is to use an agent-based modelling approach to understand and explain the dynamics of price in financial markets.

The remaining contents of this chapter are organised as follows: we first review the literature of Agent-based models. Then, an apparent weakness of agent-based models (ABMs) is discussed: they are often not calibrated to check the validation of model outputs against observed data. We firstly reviewed existing conceptualised validation approaches for ABMs, after which several well-known papers on ABMs evolving a

calibration procedure are outlined. Before describing the research objectives of this thesis, we will present recent studies that use the Genetic algorithm (GA) to calibrate ABMs, as we will apply a GA to calibrate our model. At the end of this chapter, we will illustrate the whole structure of this thesis.

1.1 Background of Agent-Based Modelling

Agent-based modelling is a research methodology that starts from the agent perspective and relies on computational tools seeking to understand how the systems work. Instead of merely describing the global phenomenon of the system, ABMs can generate the phenomenon from the actions and interactions of the multi-agent system. This bottom-up nature of ABMs makes it particularly suitable for analysing complex systems and emergent events in biology, traffic, economics and others.

Klügl and Bazzan (2012) pointed out that three essential elements are required to be explicitly considered for creating an ABM (see [5]). First, the set of agents is the most distinctive element. Agents can be broadly defined as representing physical, biological and institutional entities or specifically referred to as investors, traders, brokers and financial institutions in the financial market. The second element is the specification of the interaction in groups of agents since these interactions take account for producing the global outcome. Last comes the simulated environment containing all other elements.

Financial markets are an excellent playground for agent-based modelling with the following features: (1) Agents can have more straightforward objectives since the information collection and price issues are more likely to be sharper in financial settings; (2) Financial data is easily achievable at many different time scales from annual to minute by minute; (3) The experimental financial markets are continuously developed under carefully controlled environments, which can be compared with agent-based experiments; (4) ABMs are pretty suitable for starting from the bottom up with simple adaptive or learning agents. In particular, heterogeneous agents with adaptive or learning strategies are widely used to build rational models. For more complicated models, agents can be designed to learn from the interactions in a closed environment and change their strategies over time in response to past performance.

A simplified conception of a financial market includes a set of market participants, a trading mechanism, and a set of securities, and an agent-based model has a similar structure, including a set of agents, topology and an environment. On 6th May 2010, the U.S. stock market experienced the largest one-day point decline in the index history that the Dow Jones Industrial Average fell about 9% within a few minutes. This so-called 'Flash Crash' started from the sale of approximately $4.1 billion worth of E-mini S&P 500 futures contract through automated execution algorithms, which subsequently triggered a massive amount of transactions across the futures and equities markets, resulting in complete evaporation of liquidity in the market. Many works of the literature on the event of 6th May 2010 and other less significant flash crashes suggested that the markets should be divided into subcategories of traders and the combination of trading styles responsible for these events. In [6], Paddrik *et al.* (2012) investigated the characteristics of agents regarding their trading speed and order placement in the limit order book and summarised six categories of trader types: (1)fundamental buyers and sellers, (2) market makers, (3) opportunistic, (4) high frequency and (5) small traders. Their model consisting of these trader classes is fundamental to reproducing a market phenomenon and is validated against empirical features of real data in price returns and volatility.

Many empirical studies have contributed to building standard representative agent models. Kim and Markowitz proposed a multi-agent model which shed light on the U.S. stock market crash in 1987 that the market price decreased by more than 20%. The reason for this sudden drop is that institutional investors had widely used a computer-based dynamic hedging strategy in the years before. Since it is hard to analytically solve the model containing interactions of many investors following the same strategy, Kim and Markowitz used Monte Carlo simulations of a complicated model of price formation in an artificial financial market to investigate the destabilising potential of dynamic hedging strategies.

After Kim and Markowitz, the literature moved on to another question: analysis of asymmetric information in groups of agents, and Bayesian learning methods provide a rigorous mathematical approach to deal with models with large classes of asymmetric information. Fontnouvelle (2000) introduced a noisy rational expectations model,

where asset trading is affected by the costs of information acquisition and expectation formation (see [7]). Agents rely on an expectation strategy to decide how much information to acquire due to the information costs. Consequently, simulated time series of return volatility and trading volume shows a co-persistence, as we can find in actual financial data. Different from Fontnouvelle's work, Routledge (2001) provided a financial learning model with a genetic algorithm that agents need to learn how to make an inference about a signal observed from a market-clearing price and make a judgement about whether a signal is worth acquiring or not (see [8]). The main interest of these two authors is to investigate whether or not learning dynamics can converge to a time-invariant equilibrium with rational expectations.

Apart from contributing to different variants of adaptive learning mechanisms in ABMs, authors showed an interest in how much intelligence was needed for computational agents to generate the results they were seeing in real financial markets. Becker (1962) first proved that agents who have a random choice behaviour subject to a budget constraint could contribute to several essential characteristics of economics, for instance, downward-trend demand functions and upward-trend supply functions (see [9]). He also pointed out that researchers should not impute all observed irrationality of agents to markets or, by contrast, impute all rationality of markets to their candidates. Meanwhile, Smith (1962) stated that an approximate economic equilibrium can be obtained as market outcomes without the need of Walrasian tatonnement, which is an iterative auction process where agents calculate their demand for goods at all choices of price then submit them to a centre auctioneer (see [10]). Also, Gode and Sunder (1993) ran market experiments in which agents randomly issue bids and offers within a pre-defined range (see [11]). A double auction market is provided, and agents are restricted to their budget. The result shows no convergence of market price when agents are not subject to budget constraints. By contrast, the market price exhibits a calmer series close to equilibrium in an experimental market with human traders. This paper points out a critical message for building artificial markets: researchers need to distinguish whether the features come from agents' learning and adaption or the market structure itself. In [12], Bartolozzi (2010) introduced a novel multi-agent model characterised by a realistic order book-keeping as a tool for the study of activity of a double auction market at microscopic time scales. Their model relied on a few basic

assumptions related to agent's strategic behaviour. Among the most important ones, the order submission process makes use of a stochastic variable, the market sentiment relating to both the public and private information. The result indicated that a large part of the dynamics of the stock market at very short time scales can be explained without the requiring any particular rational approach from agent prospective. Moreover, the results confirmed that large fluctuations in price are more likely to related to a temporary lack of liquidity in the limit order book rather than to large volume transactions.

Other artificial markets include that Hansen and Singleton (1983) presented a representative consumer model to investigate the time-series behaviour of asset returns and aggregate consumption (see [13]). Chen (2001) and Yeh (2002) constructed a model that uses genetic programming to evolve a population of agents learning over time (see [14] and [15]). Tests were implemented to examine the aggregate behaviour of the time series generated by the model. They found that some series can not reject the efficient market hypothesis or the rational expectations hypothesis, proving that ABMs can empirically replicate some well-known economic behaviour. Moreover, Raberto *et al.* (2001) proposed an ABM where the price formation is based on one single asset tradings among heterogeneous agents using a realistic trading mechanism (see [16]). The model is also a flexible computational, experimental facility to accommodate various learning devices.

The adaptive belief dynamic is an alternative research topic in learning models. Agents are assumed to switch between predictors based on their past performance with adaptive belief dynamics. The best-known work in financial markets is definitely the Santa Fe Artificial Stock Market, whose design philosophy origins from understanding the influence of agent interaction and learning dynamics in groups under a financial setting (see [17], [18] and [19]). The model consists of two classifiers of agents, which are simple chartist and fundamentalist rules. Certain combinations of classifiers dominate predictions of future returns. On the other hand, classifiers and predictions are subjected to genetic operations, such as selection, cross over and mutation. In particular, the successful combinations of classifiers are more likely to be maintained, while the bad ones should be ignored in favour of better ones. The primary funding of the

early work at the Santa Fe Artificial Stock Markets was that the frequency of activation of the genetic operations significantly influences the dominance of either chartist or fundamentalist classifiers. Moreover, the chartists are more likely to dominate the market with more frequent activation of the genetic operations. Whereas Lebaron *et al.* (1999) showed that the Santa Fe model could reproduce some empirical fact merging in real-world markets, for instance, leptokurtosis of returns and correlation between the order volume and volatility (see [18]).

Another branch of literature proposed to discuss the diversity of behavioural variants. Landes and Loistl (1992) may show the first attempt using the agent-based approach with stochastic features in financial markets (see [20]). They explicitly considered a model built with the self-organising micro-structure of the stock exchange, where the process of offers, trades and the adaptive expectation of agents are modelled with the help of stochastic jump processes. The model is able to produce continuous quotations of asset prices at auction type stock exchanges and provides the possibility to simulate the variability and richness of capital market scenarios explicitly. Youssefmir *et al.* (1998) presented a model consisting of heterogeneous agents involved in an asset market (see [21]). They showed that the bubbles occur when speculative trends dominate over fundamental beliefs, which leads to asset prices growth far away from their fundamental value. The market is increasingly susceptible to this abnormal growth of any exogenous stock and finally precipitate a crush. Later work includes Farmer and Joshi, who considered a model that the price formation is based on market makers. Several popular trading strategies are used to study the price dynamics and how these strategies can amplify noise in the structure of prices, eventually causing phenomena such as excess and clustered volatility (see [22]). Similarly, Carvalho (2001) provided a simplified variant of Farmer's model, where they approached the uncertainty of agents' behaviour by allocating each agent with an unconditional probability of being activated at each time step (see [23]. The model illustrated that Pareto tailed return exists even if value investors are the only type of strategies used by the participants in the market. Aoki (2002) made another highly relevant contribution by using a jump Markov process to model entries, exits and switching of trading rule for market participation (see [24]). He showed that the fraction of the two most significant subgroups of agents with two trading rules is approximately 95

%, supporting many models of speculative dynamics consisting of two trader groups with a theoretical rationale.

In the real markets, the participants are usually classified as either the chartists or fundamentalists by their trading strategies, and many works of literature are interested in modelling the interaction of these two types of traders. The first paper on this subject can be found back in the 1950s, Baumol (1957) (see [25]) analytically discussed the destabilising potential of chartist strategies in exchange markets. Due to the lack of rationality of agents' behaviours, the topic about the chartists and the fundamentalists was dropped. However, some meaningful contributions to this strand of literature could still be found in the 1970s and 1980s. Zeeman (1973) explained the unstable behaviour of stock exchanges based on catastrophe theory (see [26] and [27]). While Beja and Goldman found that an abundance of speculative activity significantly influences stock prices' dynamics. For instance, in [28], Day and Huang (1990) proposed an excess demand and price adjustment model to answer a question: Can market participants' behaviours be used to generate observable features in bull and bear markets?

Nevertheless, the chartists-fundamentalists interacting model is also applied in foreign exchange markets. Frankel and Froot (1986) combined a standard monetary model with a chartist-fundamentalist interacting system to form expectations in open macroeconomics (see [29] and [30]). Their model is attempted to give a possible explanation of the dollar bubble that happened between 1980 and 1985. They found that the motion of a self-reinforcing interplay between forecasts and actual development can cause a price to deviate from the fundamental value. Some agents will switch their strategies from the fundamentalist to the chartist after observing the initial deviation between price and fundamental value. Furthermore, the prices have less pressure to revert to their fundamental anchor values if more market participants favour the chartist behaviour.

DeGrauwe *et al.* (1993) conducted a vital variant of Frankel and Froot's work that a monetary model is constructed with the interactions between the fundamentalists and the chartists, where a similar dynamic as Frankel and Froot's model can produce a chaotic behaviour of exchange rates [31]. Their model is capable of producing two well-known empirical facts observed in foreign exchange markets: (1) A unit root exists in the behaviour of speculative prices; (2) The forward premium is biased for predicting

the future change in the exchange rate. However, they also showed that it is difficult to distinguish their chaotic dynamics from a pure random walk process. Another literature in foreign exchange markets includes that LeBaron (1998) performed tests on several foreign exchange series generated by different technique trading rules. The results show that these series simulated regime shifts and persistent trends can not capture some aspects of the actual series [32].

When it comes to the twenty-first century, the development of technology has resulted in trading using sophisticated algorithms (programmed systems) to automate all or some part of the trade process, known as algorithmic trading. Algorithmic trading is growing rapidly across all types of financial instruments, accounting for over 73% of U.S. equity volumes in 2011 [33]. Indeed, high-frequency trading is a form of algorithmic trading characterised by high turnover and high order-trade ratios. In [34], Cartea *et al.* (2016) proposed a model where an algorithmic trader takes a view on the distribution of prices at a future. The paper contributes to a growing literature in many aspects and a new direction. To the authors' knowledge, it is the first paper that incorporates information and dynamic learning in an algorithmic trading problem through a stochastic control method. Agents learn from a collection of assets (i.e. use the current prices to update the prior) and trades in all or a subset of them with a directional strategy. Moreover, they showed how trade is made by both a limit and a market order.

Guilbaud and Pham (2013) have followed a simlar path where they studied optimal market-making policies in the limit order book, whose objective is to maximise the expected utility from revenue for the market maker [35]. The model can be seen as a mixed regime-switching regular problem based on a quasi-variation system using dynamic programming methods. A calibration procedure is provided by estimating the transition matrix and intensity parameters for both the bid-ask spread and the execution of limit orders modelled by Cox processes. The authors solved the problem with an explicit system of simple equations containing only the inventory and spread variables and illustrated the influence and profit when considering the execution priority between limit orders and market orders.

Cartea and Jaimungal (2013) developed a hidden Markov model to answer the following two questions [36]: (1) How does the intra-day dynamics of the stock market

change? (2) How can investors use intra-day information to develop trading strategies at high frequencies? In particular, The model is capable of earning profit from the bid-ask spread by submitting limit orders and is the evidence for high-frequency traders making a profit from liquidity incentives or liquidity rebates. Later, Cartea *et al.* (2018) developed a trading strategy with limit and market orders in a multi-asset economy where the assets are correlated and structurally dependent [37]. The structural dependence is established on a mid-price process following a multivariate reflected Brownian motion dominated by the asset's bid-ask spread on the closure of a no-arbitrage region. An optimal control problem is proposed for investors who take positions in these assets, while the optimal strategy depends on the distance of the vector of mid-prices to no-arbitrage bounds. Their numerical results indicate that market orders play an essential role in the strategy's success for statistical arbitrages and inventory management.

Furthermore, Leal *et al.* (2016) paid attention to the impact of high-frequency trading on asset price dynamics by building an agent-based model of a limit order book market, where heterogeneous high-frequency traders interact with low-frequency ones (see [38]). The model is able to replicate the main stylised facts of the financial market: (1) zero auto-correlation detected in price returns and the auto-correlation functions of both absolute and squared returns display a slow decaying pattern (See [39]), (2) existence of fat tails in the distribution of price returns. Their model explained the emergence of flash crashes with the following facts. High-frequency traders cause periods of high illiquidity represented by large bid-ask spreads and the synchronisation of their orders on the sell side of the limit order book. Additionally, low-frequency traders concentrate on the buy-side of the book. Moreover, they found that the cancellation of high-frequency traders plays an essential role in shaping asset price volatility and the frequency as well as the duration of flash crashes. However, they also stated that high-frequency traders speed up the recovery of market price after a crash.

1.2 Validation and Calibration

We have reviewed many excellent works of the literature based on Agent-based Modelling and noticed an apparent weakness: ABMs are not often calibrated to check the

validation of model outputs. This weakness can be credited to two contrasting aspects of ABMs that a variety of parameters included in the model might fit any feature of real observations. On the other hand, it causes a significant difficulty to access an analytical expression for model evaluation. Therefore some knowledge about the empirical validation of ABMs is discussed in this section.

In [40], Windrum *et al.* (2007) presented an influential survey in empirical validation of ABMs. They stated that very different approaches exist in how agent-based modellers conduct empirical validation. This methodological heterogeneity in ABMs is usually caused by two main factors: (1) highly non-linear system, i.e. systems with stochastic dynamics, non-trivial interaction in groups of agents and feedback from the micro and macro level; and (2) different structural content of ABMs. Figure 1.1 shows a novel taxonomy of ABMs capturing this diversity in four dimensions. In particular, the nature of the objects are the stylised facts or empirically observed facts that the model intends to explain. Examples range from qualitative information used in fundamental analysis like periodic reports to quantitative objects such as statistical properties of stock log-return, like auto-correlation patterns. Another notable distinction exists between ABMs that seek to investigate a single phenomenon, such as the return distribution of a single asset and those that jointly investigate multiple phenomena, for instance, the return distribution of a single asset together with monetary policies and agent's trading strategy. Moreover, ABMs may concern the transient or the long-run impact of the research object and can investigate from the micro-level such as individual behaviours, to the macro-level, like herding effects in the system.

The goal of the analysis can differ between ABMs that tend to deal with in-sample data with a prime purpose to replicate statistical properties of historical data and those that aim to answer control-related problems, make predictions or address policy implications with out-of-sample exercises. The modelling assumption is the most important dimension that creates a diversity of ABMs in parameters, decision rules and interaction structures. Last, to access the properties of an ABM, a detailed sensitivity analysis is needed to explore how the output depends on micro-macro parameters, initial conditions, and across-run variability included by the stochastic element, such as random individual behaviours.

These four critical dimensions firmly instruct the choice of the empirical validation

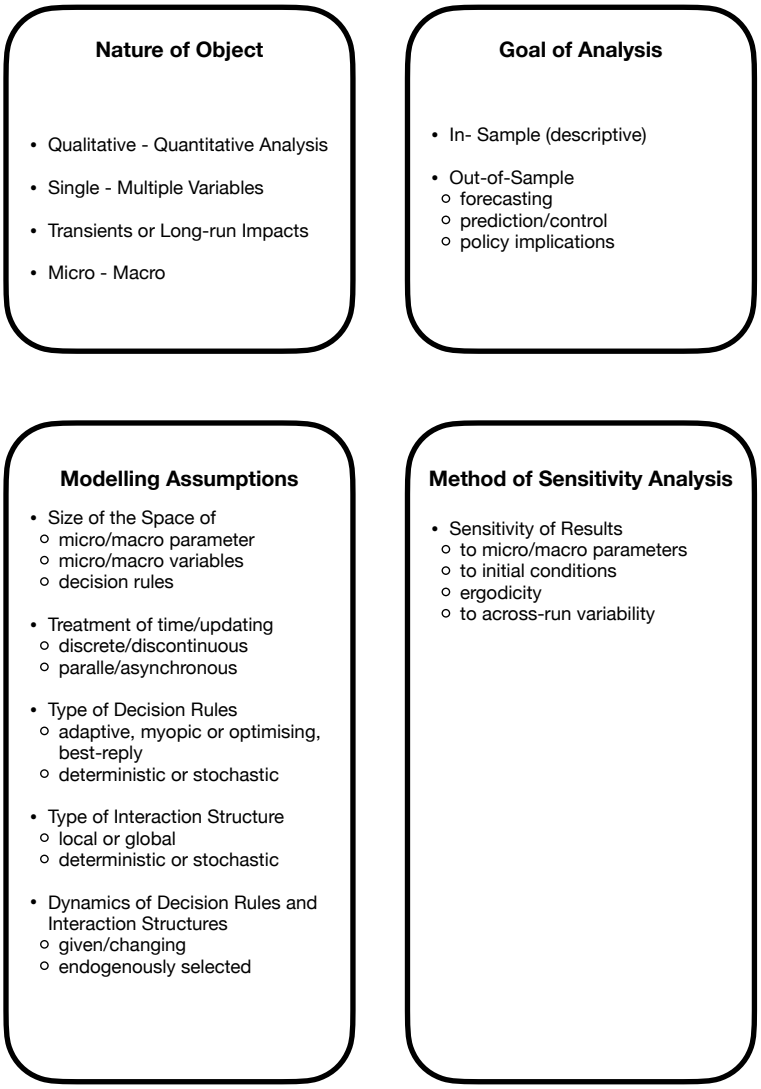


Figure 1.1: Taxonomy of dimensions of heterogeneity in ABMs introduced by Windrum

procedure. The focus on the nature of the object determines the type of data required for empirical validation, the statistical procedures and the ability to give testable empirical implications. At the same time, sensitivity analysis is a prerequisite for empirical validation, which has significant implications for the universality of the simulation output. Either the analysis is an in-sample experiment or intends to make an out-of-sample prediction, it provides different data collection and analysis approaches. In addition, out-of-sample analysis needs a calibration procedure for parameters and initial conditions.

Besides presenting a novel taxonomy of dimensions of heterogeneity in ABMs, Windrum *et al.* (2007) discussed three major approaches to agent-based empirical

validation. First, the indirect calibration approach requires modellers to identify a set of stylised facts that they are interested in reproducing and explaining with an ABM. The empirical calibration procedure seeks to make the micro-level description of the model as close as possible to the empirical and experimental evidence about micro-level behaviours and interactions. If the model output turns out to be non-ergodic, the restrictions of the space of parameters and the initial conditions are used according to the empirical evidence on stylised facts.

Second, the Werker-Brenner (2004) approach is a three-step procedure for empirical calibration (see [41]), which firstly uses existing empirical knowledge to calibrate initial conditions and the range of model parameters. Then the empirical validation of the outputs is conducted for each model specification derived from calibration. In particular, each model specification will be accepted with a likelihood relative to the percentage of theoretical realisations compatible with each empirical realisation. Last, a further round of calibration is implemented using the surviving set of models or expert testimony from historians.

Like the calibration approaches introduced above, the history-friendly approach provides a solution to the problem of over-parameterisation, which also pays more attention to the trade-off with isolating some causal mechanisms in the ABMs. This approach calibrates a model using historical case studies of the industry to model parameters, interactions among agents and agent's decision rules.

Another literature review of existing calibration and validation techniques between 1990 and 2016 was done by Fagiolo *et al.* (2017) (see [42]) that they sketched a simple theoretical framework that conceptualised existing validation approaches by answering the following four questions

- how the parameters and initial conditions are chosen?
- how summary statistics are selected and connected to characterise model's behaviour?
- how the fitness between simulated and real-observed data is measured?
- how the space of initial conditions and parameter values is investigated?

Meanwhile, they pointed out that the validation of agent-based modelling will never tell whether a model accurately describes the complex, unknown and non-understandable

real-world data generating process. However, it should eventually allow researchers to understand whether a model is a bad description of it.

A significant difficulty in calibrating the ABMs is that the complex micro-level interactions and the presence of nonlinearities (even in the simplest models) do not allow a modeller to obtain a closed-form solution of the likelihood function and the moments' conditions. Therefore, the indirect inference introduced by Gourieroux *et al.* (1993) is adapted that allows one to make inferences about the parameters of a model through simulation methods (see [43]). Also, an approach called Method of Simulated Moments is widely employed in most financial and economic ABMs, especially when the moment function is entirely unknown. It aims to minimise some statistics (moments) distances between data simulated by ABMs and data observed in the real world. While most of the calibration works follow a frequentist approach, the Bayesian approach for calibrating ABMs have been introduced in Grazzini *et al.* (2017) (see [44]) that reduce the range of choices of moments, auxiliary models and metrics evaluating the distance between the real and simulated time series. On the other hand, the Bayesian approach could be more asymptotically efficient since it uses the information over the while distribution of data, not only some specific moments.

The validation of ABMs consists of a lot of inter-related issues and concepts. Input validation can focus on the aspect of (1) tests of some behavioural assumptions typically included in ABMs, (2) selection of the initial model conditions, and (3) exploration of the parameter space (see [42]). By contrast, output validation evaluates to what extent the outcome of a simulated model fits the real-world observations, and the baseline evaluation process naturally embedded in most of ABMs emphases on replicating a set of stylised facts. Furthermore, some sophisticated statistical techniques have been designed to meet the requirements of output validation. For instance, Marks (2007) implemented three similarity measures, the Kullback-Leibler, the State Similarity measure and the Generalised Hartley Metric, to analyse and validate an ABM in a general framework (see [45] and [46]).

Although many empirical studies in ABMs can give descriptive output validation or input validation, few of them address the field of predictive output validation. Besides, most calibration techniques are based on the trial and error approach, which

may mismatch the simulated and real-time series. An important exception is a calibration procedure for validation presented by Recchioni *et al.* (2015), which is based on a famous ABM in the financial area introduced by Brock and Hommes (1998) (see [47] and [48]). They proposed a non-linear constrained optimisation problem for the calibration process, which can be solved numerically using a gradient-based method. This calibration procedure, combined with a simplified version of the Brock and Hommes model, can reproduce the daily price series of four different indices in stock markets: the S&P 500, the Euro Stoxx 50, the Nikkei 225 and the CSI 300. The calibration of the model parameters can extract some information on the micro-market behaviour, and values of the parameters show the differences and the similarities in the behaviour of agents operating in these four markets. Therefore, this calibration framework can answer questions about risk aversion, imitative behaviour, and agents' strategies in the financial markets. Another significant contribution made by Alfarano *et al.* (2006) is that they analytically solved a return probability distribution of an ABM and calibrated the model parameters by the maximum likelihood method (see [49]).

1.3 Genetic Algorithms

A problem in agent-based modelling is that ABMs consisting of a large set of parameters or stochastic dynamics are hard to get an explicit formula for simulated output. The calibration of the parameters by proposing an optimisation problem is blocked due to the derivatives of the considered parameter being inaccessible. Therefore, the Genetic Algorithm (GA) inspired by the natural selection procedure attracts many modellers interest. The main features of the algorithm are mutation, crossover and selection operators, which are supposed to mimic those we have observed as part of natural selection in the real world. Meanwhile, it has been widely used to produce high-quality solutions to optimisation and search problems in various fields.

Routledge (2001) [8] employed GA in an ABM that agents decide whether a signal of the market price is worth acquiring, make an inference about the signal and learn how to use this signal. They provided some examples showing that GA does and does not converge to the rational expectations equilibrium, and the behaviour is mainly

affected by two factors: the rate of or mutation in the GA and the size of the risky-asset supply noise in the economy. Rogers and Von Tessin (2004) presented an ABM of a financial market, which is calibrated using a multi-objective GA (see [50]). Their model evolves a population of Pareto-optimal parameters sets, in which a single candidate can be selected as a final tuned parameter set without requiring any explicit prior weighting of criteria. Another classic model is proposed by Farmer and Joshi (2002) [22] that the calibration process is conducted by using a Nelder-Mead simplex algorithm and a GA [51]. Both methods can fill the gap of parameters estimation in the ABMs field and are more robust in noisy environments.

1.4 Research Objectives

The objective of this thesis is to model the stock market via the agent-based approach. We will provide a framework with elements, such as the trading mechanism, communication network system, and ordered events for daily operation, which are essential to creating an artificial stock market. The basic agent-based model is built on the assumption that agents randomly place a bid or ask order in the market, who are called the noisy trader with no trading objective. Meanwhile, an expectation feedback system is embedded in the model, cooperating with the endogenous imitation mechanism. We will show that this basic model can reproduce a set of stylised empirical facts emerging in real financial price series, for example, zero auto-correlation of log return and volatility clustering. The calibration of the model is implemented by proposing an optimisation problem seeking to minimise the linear combinations of moment difference between observed and simulated log-returns, and we use a genetic algorithm to find the region close to the optimal parameter rather than a single value due to the stochasticity and complexity of the simulation outputs.

An extension of the basic agent-based framework is based on defining agents with different trading behaviours, such as holding a fixed belief about the asset return or following the trend of the asset return. Agents are aiming to maximise their utility function through these trading behaviours. Therefore, we will make use of stochastic control theory and the Bayesian learning method to find the optimal strategy of agents under the utility maximisation with different trading behaviours.

Under the agent-based framework, we are able to create a mixed market that has noisy and trend-following agents and investigate such research questions as:

- Under what setting can we produce a stylised market with a deterministic trend?
- Under what settings the market will crash?
- Can we identify the pattern in the trending following case?
- Are there any early warning signs that market instability is approaching?

1.5 Thesis Structure

This thesis uses agent-based modelling approaches combined with heterogeneous agents to produce an artificial financial market. The basic framework of modelling was presented initially by Tedeschi *et al.* (2009) [52], which consists of a characteristic trading mechanism, an endogenous mechanism of imitation and noisy agents. We extend their work by equipping agents with optimal portfolio strategies introduced by Bismuth *et al.* (2019) [53] that agents are capable of learning from the historical price to adjust their risky asset proportion of the portfolio.

In Chapter 2, we begin with mathematically formulating the behaviour of the noisy agent, such as the future price expectation and the order placement. Following Tedeschi *et al.* (2009) [52], an endogenous mechanism of imitation is introduced, allowing us to create a herding effect in the system that each agent can be imitated by others with a probability proportional to his wealth. We verify the presence of a set of stylised empirical facts in the model, for example, a heavy tail distribution, zero auto-correlation in log-return and volatility clustering. We also contribute to showing a parameter sensitivity analysis, which helps choose the initial condition and a further model calibration.

Chapter 3 extends the agent-based model by presenting a utility-maximisation framework of Bismuth *et al.* (2019) [53] for trend-following agents who use the Bayesian learning method to track the asset return. The portfolio choice problem is proposed with two cases that agents have the constant relative risk aversion (CRRA) or the absolute risk aversion (CARA) utility function. Using the ABM with agents having optimal portfolio strategy, we contribute to illustrating how a market can be

set up with a deterministic trend and what leads to a crash in the system. Meanwhile, the parameter sensitivity analysis is presented to find the suitable initial condition and constraint of parameters for a traceable and reproducible simulation to exist.

Chapter 4 mainly focus on calibrating the ABM against real-world data. We start from using a gradient-based method and a jump detection approach to fit an ABM. Although these two methods fail to give a reliable calibration, they provide us with a hint to use a linear combination of the log-return moments to capture the main features of data. Next, we show that the genetic algorithm (GA) is more robust in calibrating stochastic outputs. For predictive purposes, we apply the GA calibration result against the real Cryptocurrency data to predict the future empirical distribution of log-returns and calculate risk metrics, such as Value at Risk (VaR) and Conditional Value at Risk (CVaR).

We conclude in Chapter 5.1 and discuss some future potential work and research to undertake in the future.

In the Appendices, proofs of Propositions introduced in Chapter 3 and a published paper in the early stage of my PhD study are given. The paper uses simultaneous Bayesian modelling to analyse data on the size of lymphedema occurring in the arms of breast cancer patients after breast cancer surgery (as the longitudinal data) and the time interval for disease progression (as the time-to-event occurrence). A model based on a multivariate skew t distribution is proven to be more efficient than the multivariate normal distribution by comparing different distributional assumptions for residuals and random effects.

Chapter 2

An Agent-Based Model for Stock Market

In this chapter, we follow Tedeschi *et al.* (2009) [52] that introduce an Agent-based model (ABM) with heterogeneous agents in stock market. The ABM is built on the assumption that agents randomly place a bid or ask order in the market, who are called the noisy trader with no trading objective. The component of an ABM includes elements, such as the trading mechanism, communication network system, and ordered events for daily operation, which are essential to creating an artificial stock market. For the communication network system, an endogenous mechanism of imitation via a preferential attachment rule [54] is applied, such that the model is able to demonstrate how an expectation feedback system can amplify fluctuations in the market price of the asset. Although it is a toy model that is built upon some endogenous rules whose purpose is not to reproduce agent behaviours in the financial markets exactly, it allows us to verify the presence of a set of stylised empirical facts in the model, for example, a heavy tail distribution, zero auto-correlation in log-return and volatility clustering.

In the following parts, we begin by introducing the notation of the model. Then, we mathematically formulate the noisy agent's behaviour, such as the future price expectation and the order placement. Following Tedeschi *et al.* (2009), a pn endogenous mechanism of imitation is presented, allowing us to create a herding effect in the system that each agent can be imitated by others with a probability proportional to his wealth. Next, we display some simulation results to verify the presence of stylised empirical facts. Finally, we also contribute to showing a parameter sensitivity analysis,

which helps choose the initial condition and a further model calibration.

2.1 Notations for the Model

For the sake of simplicity, the model is built on a discrete-time with T periods with a total number of N^{agent} noisy agents. We denote t_k as the period, which starts from t_1 and ends with t_T ($T < \infty$). Within each period, there are L intra-day periods denoting as t_{k_l}, then t_{k_1} and t_{k_L} represents the first and the last intra-day period t_k respectively. Moreover, we define t_0 as the initial period of the model at which agents' portfolios and their networks are established.

All agents are assumed to have a random demand and supply function, and their holding cash and stocks bound the size of their orders. Credit transactions are not permitted in the system, which means that agent limit orders are limited by their available money and available shares in the portfolio. Assume that agent i has an account book, which consists of the following 6 items:

- *Cash.* C_t^i represents the total money agent i is holding at time t, which will increase/decrease by the amount of cash in a completed buy/sell transaction.

- *Limit Cash.* It represents the amount of cash occupied by limit bid orders submitted by agent i. The limit cash will decrease when limit orders are executed or cancelled.

- *Available Cash.* M_t^i represents the amount of cash that agent i can use to buy shares. The available cash will increase when shares are sold. Otherwise, the limit bid orders are cancelled. On the contrary, the available cash will decrease when shares are bought or the limit bid orders are submitted.

- *Share.* S_t^i represents the number of shares held by the agent i at time t, which will increase/decrease by the number of shares in a completed buy/sell transaction.

- *Limit Share.* It represents the number of shares occupied by limit ask orders, which will decrease when limit orders are executed or cancelled.

- *Available Share.* N_t^i represents the number of shares the agent i can sell, which will increase when bid orders are completed or limit ask orders are cancelled.

While both market and limit ask orders will decrease the available share.

Initially, all agents are provided with an amount of cash C_0 and S_0 shares. Given that P_t is denoted as the asset price at time t, then the initial wealth is defined as $W_0 = C_0 + S_0 P_0$. Meanwhile, we also define the mean price of the system as

$$P^{mean} = \frac{\sum_{i=1}^{N^{agent}} C_t^i}{\sum_{i=1}^{N^{agent}} S_t^i}, \tag{2.1}$$

which indicates the equilibrium price in the model. If no extra cash is injected into the system, this equilibrium price will stay fixed for all t.

The market limit order book collects all orders submitted by agents. Since the order is submitted within the intra-day period, then the corresponding bid and ask order price is defined by $b_{t_{k_l}}$ and $a_{t_{k_l}}$ respectively. While the volume of the order is represented by $v_{t_{k_l}}$.

2.2 Noisy Agent Description

Noisy agents are assumed to have random expectations on future price returns and a random demand function. We choose noisy agents rather than agents with sophisticated strategies since it allows one to focus on the impact of imitation on price dynamics. Moreover, the presence of noisy agents and their influences on the price movement has been well studied.

2.2.1 Agent's Future Expectation of the Price

At the beginning of each period t_k , noisy agent i will randomly assign his expectation about the price one period forward and the price they expected price at time t_{k+1} is given by

$$\hat{P}^i_{t_k,t_{k+1}} = P_{t_k} e^{\Delta t \hat{r}^i_{t_k,t_{k+1}}}, \tag{2.2}$$

where $\hat{r}^i_{t_k,t_{k+1}}$ is the agent's expectation on the spot return. We also set $\Delta t = 1$ as the smallest time step. We can calculate the spot return according to the equation

$$\hat{r}^i_{t_k,t_{k+1}} = \sigma^i_{t_k} \epsilon_{t_k}, \tag{2.3}$$

where $\sigma^i_{t_k}$ is a positive, agent specific constant and $\epsilon_{t_k} \sim N(0,1)$ is a normal distributed random variable, independently sampled at t_k. The agents will use this price to decide on their next order to submit to the market.

2.2.2 Place Orders

Since all agents have calculated their expected price, an agent is randomly selected to submit his bid or ask offer to the limit order book at each intra-day period. Given that the expected price $\hat{P}_{t_k,t_{k+1}}$ reflects the agent's beliefs on the price movement within the next period, then we can use this information to create a likely bid or ask order for the agent. For instance, if the agent believes the price will increase, we assume the agent will take a long position of the asset. By contrast, if the expected price of the agent is lower than the current price, the agent will short his position.

Let P_{t_k} represent the opening price at period t_k. Clearly, if $\hat{P}^i_{t_k,t_{k+1}} \geq P_{t_k}$, the maximum bid price for the agent i is $\hat{P}^i_{t_k,t_{k+1}}$, because he will not want to buy at a higher price than expected. Conversely, if $\hat{P}^i_{t_k,t_{k+1}} < P_{t_k}$, the agent i will bet on ask side with an ask price larger than $\hat{P}^i_{t_k,t_{k+1}}$, as he will not want to sell at a price lower than expected. Then the order price of agent i is uniformly drawn around the current price P_{t_k} and the corresponding formulas are

$$\begin{aligned} b^i_{t_{k_l}} &\sim U(P^b_{min},\, \hat{P}^i_{t_k,t_{k+1}}) \\ a^i_{t_{k_l}} &\sim U(\hat{P}^i_{t_k,\,t_{k+1}}, P^a_{max}) \\ P^b_{min} &= P_{t_k}(1-\gamma^1_{t_k}) \\ P^a_{max} &= P_{t_k}(1+\gamma^2_{t_k}), \end{aligned} \tag{2.4}$$

where $\gamma^{1,2}_{t_k}$ are random variables uniformly distributed in the interval $(0,1)$.

Order volume $v^i_{t_{k_l}}$ of agent i is determined by his order price, available money $M^i_{t_{k_l}}$ and shares $N^i_{t_{k_l}}$ in the account. The volume is randomly sampled from a uniform distribution according to the following expressions

$$v^i_{t_{k_l}} \sim \begin{cases} U\left(0,\, \frac{M^i_{t_{k_l}}}{b^i_{t_{k_l}}}\right), & \text{for bid order} \\ U\left(0,\, N^i_{t_{k_l}}\right), & \text{for ask order.} \end{cases} \tag{2.5}$$

2.3 The Communication Network

At the beginning of each period t_k, agents are assumed to join the communication network, in which agents can imitate their neighbours' expectations and share their own with others. Then agents' decisions are influenced by their interactions, which is illustrated in Figure 2.1 that the pink nodes represent agents, and the edges are the one-way directional arrow between them.

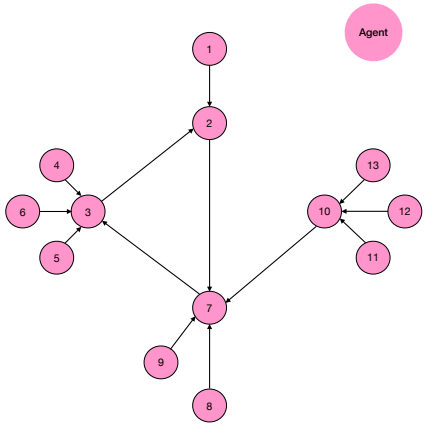


Figure 2.1: The pink nodes represent agents, and edges are the one-way connective arrows between them. Each agent can have several incoming links but only one outgoing link.

Each agent can only choose one neighbour, to whom he would send a request for advice, but he can share his opinion with unlimited imitators. The agent who has the most number of requests at each period is call the guru. Therefore, the original forecasting of agent i (Equation (2.3)) is updated as a reviewed expectation on spot return,

$$r^i_{t_k,t_{k+1}} = w\hat{r}^i_{t_k,t_{k+1}} + (1-w)\hat{r}^j_{t_k,t_{k+1}}, \tag{2.6}$$

where $w \in [0,1]$. The revised expectation $r^i_{t_k,t_{k+1}}$ consists of a linear combination of agent i and j's original expectations on spot return. Moreover, when $w = 0$, agent i completely trusts on agent j, if $w = 1$, agent i's decision is independent of the other agents.

Under a small w condition, agents can be easily affected by others' decisions. We assume that agent i who has several imitators would forecast a larger price volatility.

Let $l^i_{t_k}$ be denoted as the number of incoming links of agent i at period t_k, then his specific volatility equation can be written as

$$\sigma^i_{t_k} = A(\sigma^i_0 + \frac{l^i_{t_k}}{N^{agent}}(1-w)). \tag{2.7}$$

The parameter σ^i_0 is uniformly distributed in the interval $(0, \sigma_0)$. The effect of this volatility formation is to favour a market order over a limit order for agents who have several incoming links. It can be explained that a larger $\sigma^i_{t_k}$ leads to a more aggressive order price choice. The wider distribution of order price on the limit book makes orders more likely to be executed. Consequently, the guru is more likely to submit a market order as well as his followers. Moreover, market orders in the same direction will cause a large fluctuation in price, supporting the original expectation of the guru.

2.3.1 The Mechanism of Imitation

The mechanism of imitation is implemented by a preferential attachment rule (see [54]), in which each agent can imitate others' behaviour with a probability proportional to his wealth. This mechanism allows us to investigate how imitation affects the asset price and the distribution of agents wealth.

At the initial time, all agents are provided with an amount of cash C_0 and S_0 shares. Then the initial wealth is defined as $W_0 = C_0 + S_0P_0$. The initial network among agents is established for the first few periods whilst keeping the guru fixed. Agents have a fair chance to choose between linking with this fixed guru or randomly linking with others except the guru or himself. Under these settings, we should expect to observe that some agents will become richer than others after several periods. As a measure of agents' success, we define their fitness at period t_k relative to their peers as their wealth divided by the wealth $W^{max}_{t_k}$ of the richest agent i_{max}

$$f^i_{t_k} = \frac{W^i_{t_k}}{W^{i_{max}}_{t_k}}. \tag{2.8}$$

Agent i starts with one outgoing link to a random agent j and possibly has some incoming links from other agents. Links are rewound at the beginning of each period in the following way: each agent would cut his outgoing link in the last period with agent j and build a new link with a randomly chosen agent h with a probability

$$p = \frac{1}{1 + e^{-\beta_{t_k}(f^h_{t_k} - f^j_{t_k})}}, \tag{2.9}$$

or he will keep the existing link with probability $1 - p$.

It is noteworthy that agent i will choose fairly between agent h and agent j if they have the same wealth. Otherwise, agent i has a probability higher than 0.5 to follow the richer one between agent h and j.

The parameter β_{t_k} in Equation (2.9) represents the intensity choice of agent, which measures how much agents in the communication network trust on the information corresponding to others' performances. Moreover, β_{t_k} amplifies the effect that the agent is more likely to imitate the rich agent according to Equation (2.9). Specifically, when β_{t_k} is zero, agents act independently of each other. However, by increasing β_{t_k}, agents will be more likely to start following the guru. In the model, β_{t_k} links to the life span of the guru. The initial value of β_0 is one, and it will increase by one if the guru can survive for one more period. The value of β_{t_k} is reset to be one when a new guru occurs.

Under the condition that the guru is the most affluent agent in the model, we expect a herding effect, resulting in large fluctuations in price. Since the majority of agents trust in the guru and imitate his expectation, it causes orders in the limit order book to stack on the same side.

2.4 The Market Matching Mechanism

The matching mechanism of the limit order book follows the Euronext and the London Stock Exchange rules. Agent choices for placing a limit or market order depend on the latest asset price. Let P_t be the market price at t. The update of the new market price P_t and how agents choose their order types are described by the following rules,

1. P_t is given by the price at which a transaction occurs. If no new transaction price occurs, a proxy price for the price is used by taking the average of the lowest quoted ask price a_t^q and the highest quoted bid price b_t^q: $P_t = \frac{a_t^q + b_t^q}{2}$. If no bids and asks are listed in the book, the proxy for the price is given by the previous traded or quoted price.

2. If $b_t^i \geq a_t^q$, agent i places a market order for purchasing v_t^i shares at the current best quoted ask price a_t^q. If there is no sufficient supply at the price a_t^q, agent

i only buys the amount available on the market at the price a_t^q, then moves on to check the second best ask price, repeating the process until he has already bought v_t^i shares or $b_t^i < a_t^q$, by placing a limit order to buy the rest shares at price b_t^i. Otherwise, agent i buys v_t^i shares at the price a_t^q. If $b_t^i < a_t^q$, agent i submits a limit order to buy v_t^i shares at price b_t^i.

3. If $a_t^i \leq b_t^q$, agent i places a market order for selling v_t^i shares at the current best quoted bid price b_t^q. If the demand available on the book is sufficient, agent i can sell all v_t^i shares at the price b_t^q. Otherwise, the agent fills the available demand at the price b_t^q, then, move on to the second best bid price, iterating the process until selling all $v_t^{(i)}$ shares or $a_t^{(i)} > b_t^q$, by placing a limit order to sell the rest shares at price b_t^q. If $a_t^{(i)} > b_t^q$, agent i submits a limit order to sell $v_t^{(i)}$ shares at price $a_t^{(i)}$.

2.5 Timeline of the Events

In previous sections, we have mathematically formalised the behaviours of agents and described some essential mechanisms, which can make up an artificial order drive market. However, the event sequence is also vital to make an identical model for repeatable simulation results.

The following list displays the sequence of events in order within a period, which is iterated throughout the simulation.

1. The communication network is built in two steps.
   - For the first few periods, one agent is randomly picked as the guru during that time, and the others have a fixed probability of linking with the guru or rewind with others.
   - The preferential attachment rule is used that each agent imitates others with a probability related to others' wealth.
2. Each agent forms the original expectations on stock return using(Equation (2.3). The revised expectation from (Equation(2.6) replaces the original one for generating the expected price Equation (2.2), if the communication network is activated in the model.

3. Every agent checks his existing limited order and withdraw it if it lasts over one period. Then one agent is randomly selected to place an order (see Section 2.2.2) and market price is updated (see Section 2.4). This step will be repeated several times according to model settings.

2.6 Simulation

In the beginning, we present a pseudocode of an ABM simulation to illustrate how we can obtain outputs based on the math formulas introduced in previous sections.

Algorithm 1 The pseudocode of an ABM simulation

Require: $N^{agent},\ T,\ L\ \in \mathbb{N}^+,\quad A \in [0,1],\quad w \in (0,1)$
  $t \leftarrow 1$
  $l \leftarrow 0$
  $P_0 \leftarrow p_0$
  set up the initial portfolio account for each agent
  while $t \leq T$ **do**
    while $l \leq L$ **do**
      if $l = 0$ **then**
        establish communication network using Equation (2.8) and (2.9)
        generate agent-specfic volatilities using Equation (2.7)
        calculate agents' original expectation on spot return using Equation (2.3)
        calculate agents' reviewed expectation on spot return using Equation (2.6)
        $l \leftarrow l + 1$
      else if $1 \leq l \leq L$ **then**
        randomly select an agent
        generate this agent's expectation on future price using Equation (2.2)
        place this agent's limit order using Equation (2.4) and (2.5)
        update the new market price P_{t_l} following the rules in Section 2.4
        $l \leftarrow l + 1$
      end if
    end while
    $t \leftarrow t + 1$
  end while
  output a set of simulated price $\{P_{1_1}, P_{1_2}, \ldots, P_{T_L}\}$

In specfic, the population of noisy agents N^{agent} is set as 100. Each agent is given $S_0 = 100$ shares of stock and $C_0 = 10000$ cash at the beginning. The initial stock price is $P_0 = 100$. In Equation (2.7), the constant A is set as 0.2 and σ_0 is equal to 1.

The results presented here are the average outcome of 10 simulations with $T = 2\,000$ periods. There are $L = 300$ intra-day periods within each period, which allows each

agent to have three fair chances to submit the order.

Firstly, we will display some descriptive figures and tables of the simulation results. Then, we will focus on the statistical properties of the distribution of stock returns and the auto-correlation of market volatility. We will verify whether the stylised empirical facts (see [55]), such as a heavy tail distribution, absence of linear correlation, and volatility clustering, exist in our model and investigate how the communication network affects these facts.

Figure 2.2 shows the sample paths from simulations with $w \in \{0, 0.4, 0.8, 1\}$, where w is defined as the degree of how much agents believe on others' opinions, according to Equation (2.6). If $w = 0$, it means agents completely follow others' behaviour, while $w = 1$ means agents make decisions independently. We can directly observe that the model can produce a stationary price series without the imitation mechanism ($w = 1$). As the effect of imitation amplifies, the fluctuations of price enlarge, which can even cause the price to be tripled under the condition $w = 0$.

Table 2.1 displays some summary statistics of simulation results, which shows a positive correlation exists between the price volatility and the magnitude of imitation effect. Moreover, the more significant fluctuations increase the difference between the historical maximum and minimum asset price. It is noteworthy that the model can produce mean reversion price series under all simulations.

	$w = 0$	$w = 0.4$	$w = 0.8$	$w = 1$
mean	106.62	99.85	101.38	100.25
std	20.67	5.89	3.08	2.83
max	291.74	131.60	116.96	109.25
min	70.17	80.86	93.48	92.72
max - min	221.57	50.74	23.48	16.53

Table 2.1: Summary table of the stock price under different conditions.

The closed transaction environment determines the mean reversion character, where the amount of money and the number of shares circulating in the market is fixed. The noisy agents mainly use the standard normal distribution to form their expectations. After that, agent orders are generated using uniform distributions regarding price and volume. We expect the price to reserve the mean price with an arbitrary positive initial starting point.

In Figure 2.3, price paths are presented with different starting points under different

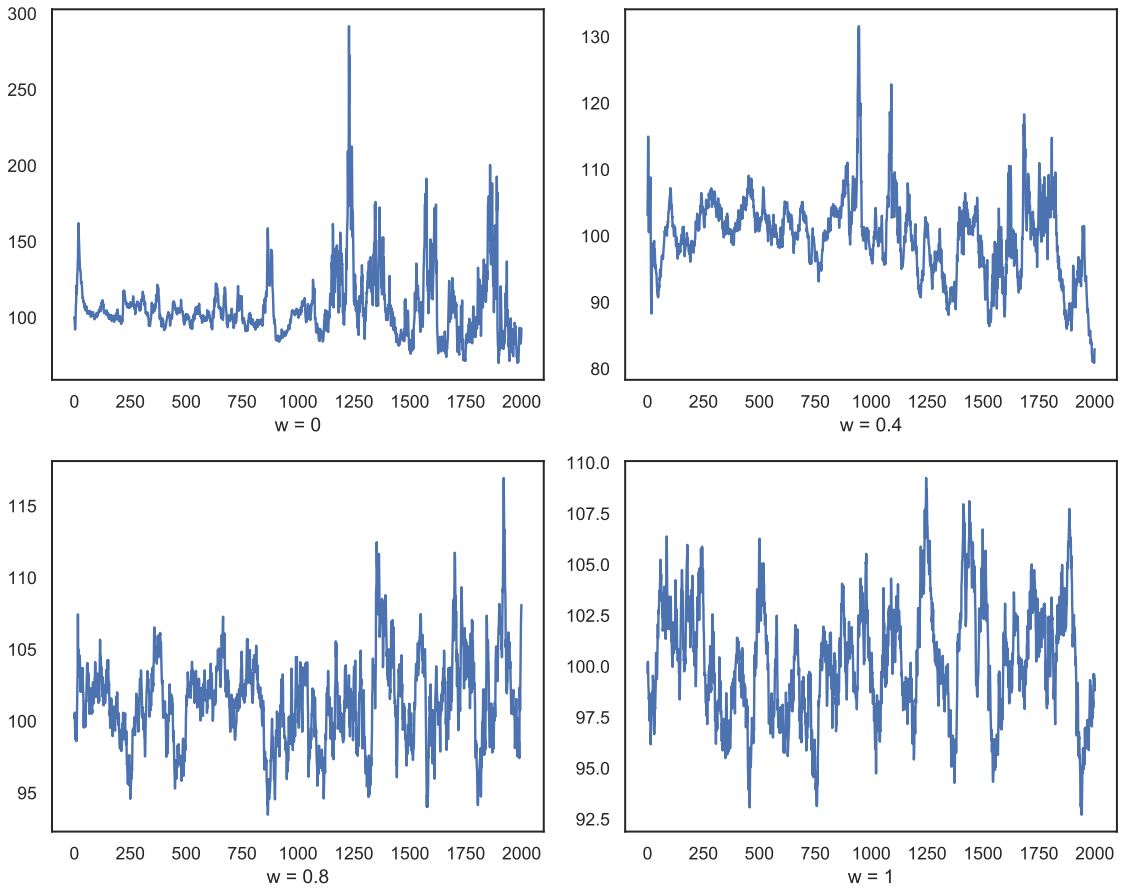


Figure 2.2: Average price time series under different conditions that how much agents believe on others' opinion (Equation 2.6). $w = 0$ means agents completely follow others' behaviour, while $w = 1$ means agents make decisions independently.

w conditions. When $P_0 > P^{mean}$, the volume of the bid order is expected to be smaller than the volume of the ask order. Since the money circulating in the system is fixed, a higher asset price leads to a lower demand capacity. Then the excess ask orders will draw the price to an equilibrium $P_0 = P^{mean}$, where the expected volume of bid order is equal to the volume of ask order. Conversely, when $P_0 < P^{mean}$, asset price will be quickly pushed up due to the excess volume of the bid order in the market.

2.6.1 Returns and Volatility

Figure 2.4 illustrates some stock return series under different w conditions. When $w = 0.8$ and $w = 1$, the return series display like a random walk. However, a larger variation of stock returns occurs when agents mainly rely on imitating others' opinions, for instance, see the upper panel of Figure 2.4. Noting that the guru is fixed at the

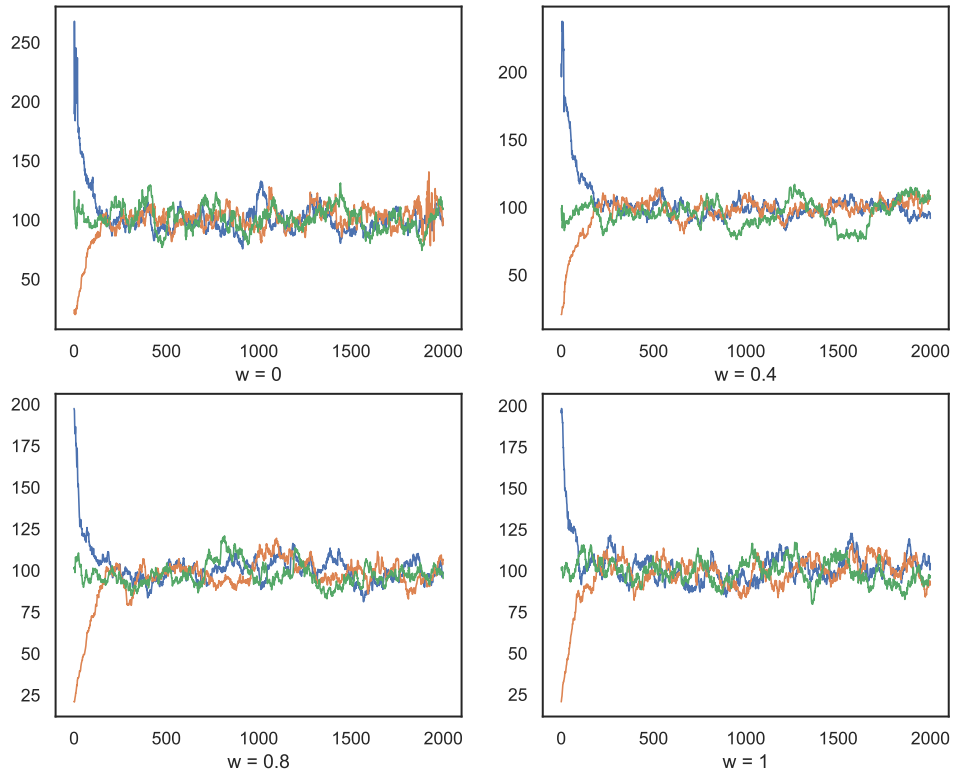


Figure 2.3: Stock paths with different starting points under different w conditions. A price equilibrium can be approached, which equals the ratio of the total amount of money to share in the market.

beginning of the simulation, and each agent has a one-half probability of linking with the guru. After the network among agents is established, some agents will become richer than others. According to Equation (2.8) and (2.9), rich agents are more likely to keep the links with their followers and then become the guru.

When a large proportion of agents follow the guru under a small w condition, the guru's expectation represents the aggregated opinion of his imitators. Whichever side the guru picks on, he and his imitators will absorb the majority of limit orders on the other side, which consequently make the price move according to the guru's will. This 'herding effect' is reflected as the humps in Figure 2.4 under the condition $w = 0$.

Figure 2.5 shows the percentage of guru's in-coming link. Due to the fixed guru setting, the percentage of guru's in-coming link fluctuates around 0.5 for the first 20 periods, which causes large excursions in stock returns under minor w conditions. After that, the preferential attachment rule is engaged, so agents have options for

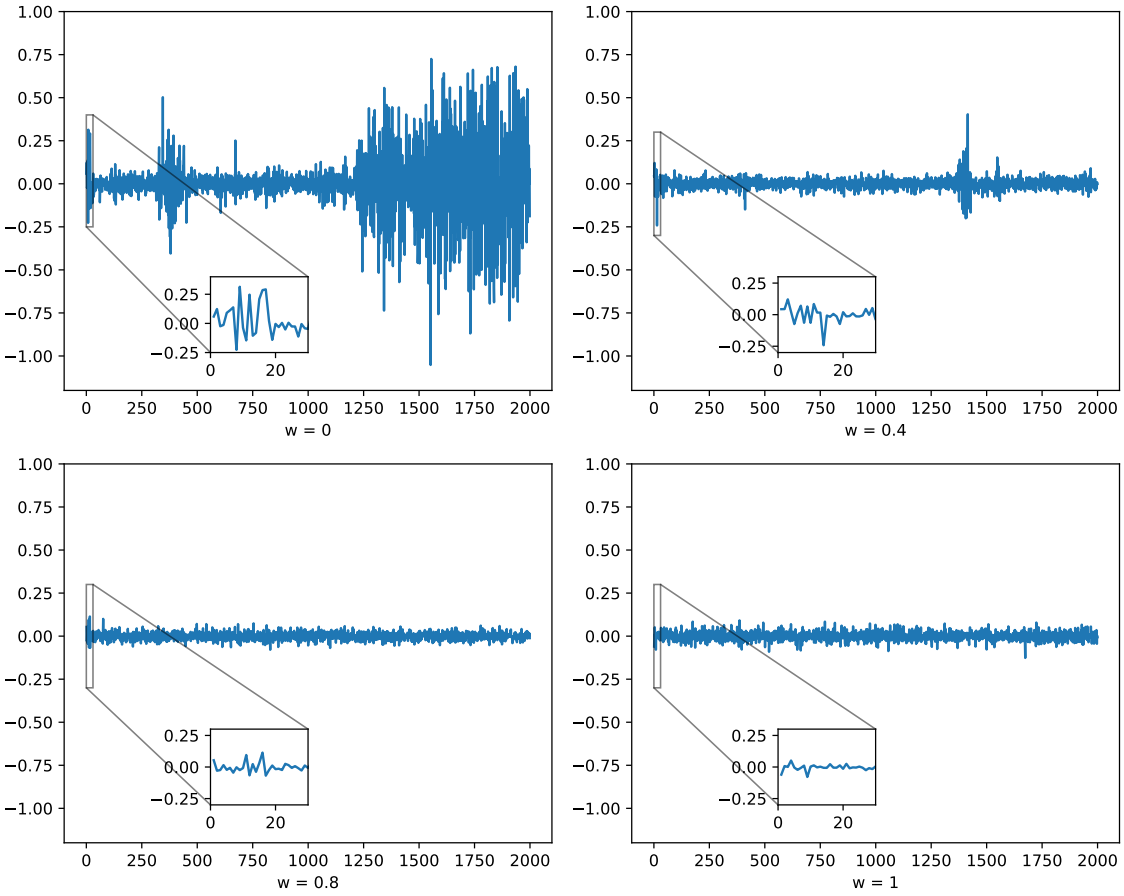


Figure 2.4: Stock return series under conditions with $w \in \{0, 0.4, 0.8, 1\}$

.

keeping the existing link or randomly rewinding with others. The agent who is more affluent in wealth has a higher probability of winning this incoming link. The agents' wealth will redistribute through the periods, and some agents, especially the guru, will gradually become more prosperous and more attractive than others. Therefore, the percentage of the guru's in-coming link keeps flat until it has an upward trend under the preferential attachment rule. Besides, the imitation mechanism will accelerate the wealth redistribution process, which causes the upward trend for guru's in-coming link to come earlier under minor w conditions. Although under the condition $w = 1$, agents make their decisions independently, the communication system still exists, and the guru will gradually have more followers.

In [55], Cont pointed out that the insufficiency of the normal distribution for modelling the asset price returns and their heavy tail character. We want to verify whether our model satisfies these empirical facts.

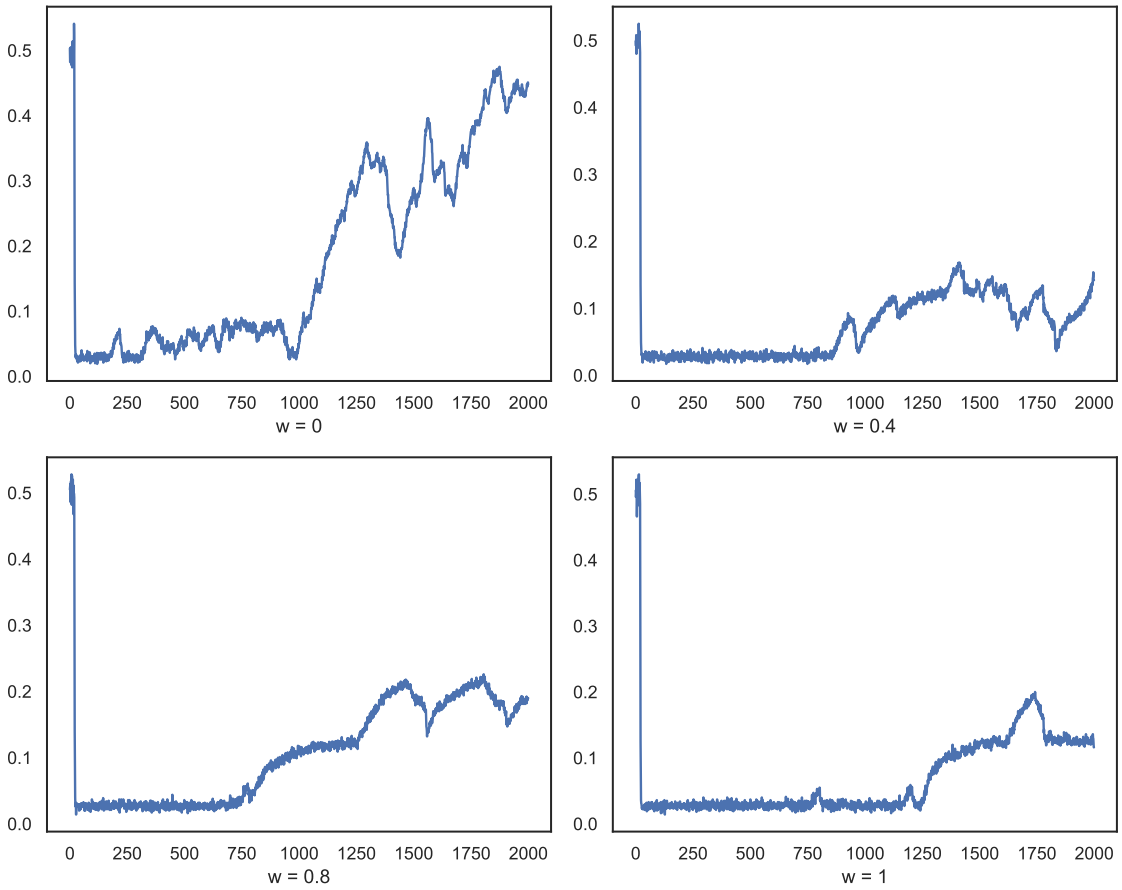


Figure 2.5: The percentage of guru's in-coming link.

	$w = 0$	$w = 0.4$	$w = 0.8$	$w = 1$
μ/σ	-0.0006	0.0071	0.0041	0.0002
Skewness	-1.00	-0.50	0.12	0.02
Kurtosis	13.42	8.94	1.71	0.10

Table 2.2: Descriptive statistics for stock returns under different w conditions.

Combining Figure 2.6 and Table 2.2, the model under four scenarios exhibits a leptokurtic, asymmetric and fat-tailed distribution for the stock return. Moreover, skewness and kurtosis of the return are amplified by increasing the imitation effect. Here, the skewness and kurtosis of return distribution are defined as

$$Skew(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \tag{2.10}$$

$$Kurt(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3 \tag{2.11}$$

where X is the return series.

The skewness quantifies the asymmetry of the probability distribution, and the

standard normal distribution has a zero skewness. The kurtosis is defined as $Kurt = 0$ for normal distribution, and a positive value indicates a fat tail, which means a slow asymptotic decay of the probability density function.

The typical values for kurtosis in real worlds are: $Kurt \approx 74$ (US/ DM exchange rate futures), $Kurt \approx 60$ (US/ Swiss Franc exchange rate futures), $Kurt \approx 16$ (S&P500 index futures) (see [56], [57], [58], [59]).

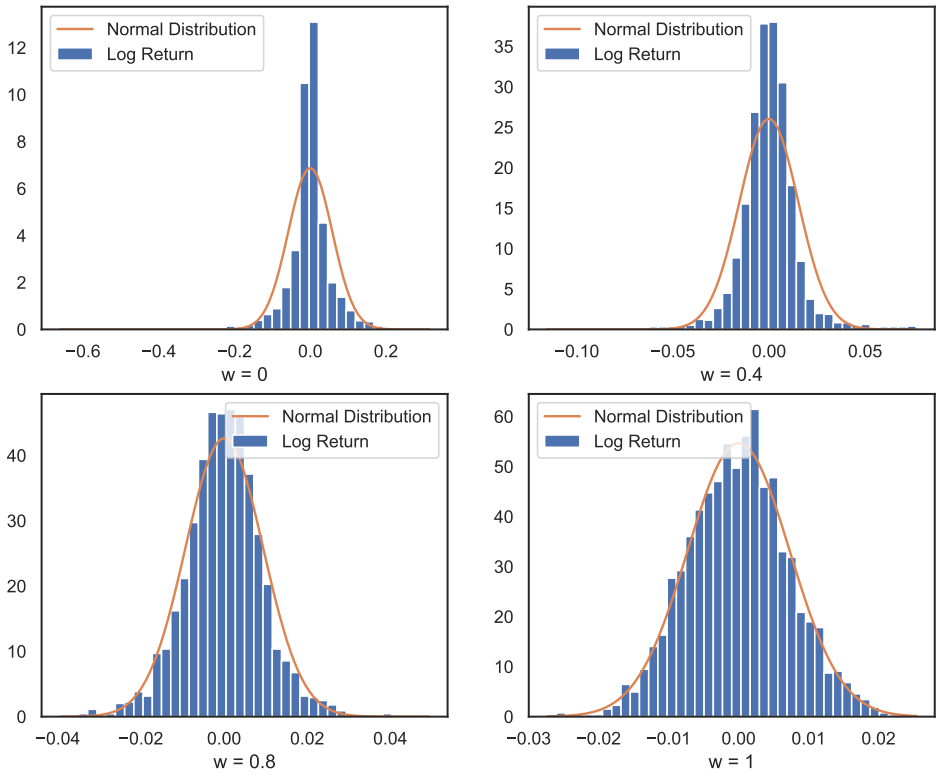


Figure 2.6: The histograms of stock returns and the corresponding density plots of normal distribution.

In [55], a well-known fact is that price movement in liquid market has no significant correlation, and it is often cited as a support for the 'efficient market hypothesis (see [60]). An intuitive explanation for the absence of auto-correlation is that if price changes exhibit significant correlation, this correlation may be used by a statistical arbitrage strategy to earn a positive expected profit. Consequently, it will reduce correlations except for short time scales, representing the market's time to react to new information.

Figure 2.7 shows the auto-correlation of stock returns under different w conditions, which rapidly decay to zero in a few lags. Although zero auto-correlation in return can provide 'random walk' models of price with some empirical support, which assumes the increment of price is independent, the absence of auto-correlation does not imply the independent increment in price. While the independent increment implies any no-linear function of returns will produce non-correlation series.

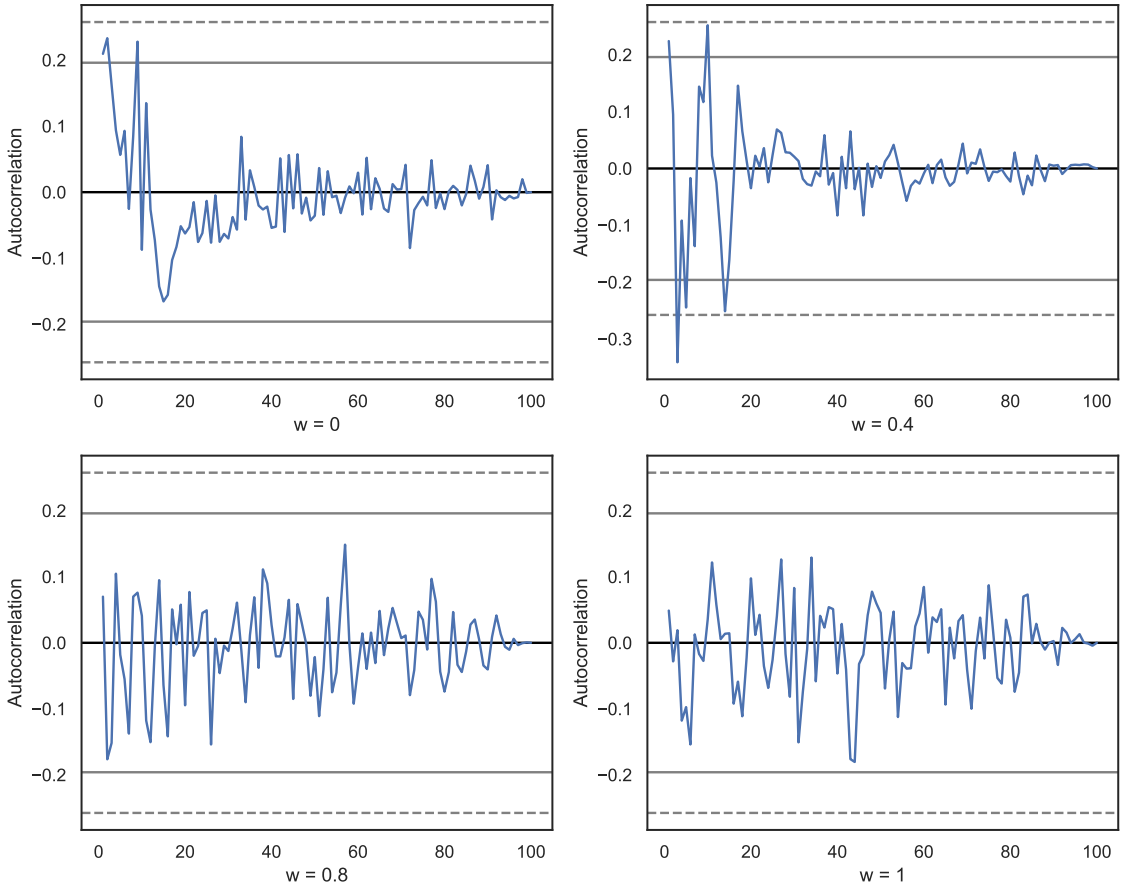


Figure 2.7: Auto-correlations of stock returns under different w conditions.

Figure 2.8 displays the absolute auto-correlation of absolute stock returns, which exhibits significant positive auto-correlation for the first few lags under $w = 0$ and $w = 0.4$ conditions. This quantitative signature supports the existence of a well-known phenomenon called 'volatility clustering', which is interpreted as large price variations are more likely to be followed by large price variations.

Recalling Figure 2.2 and 2.4, large price variations are caused by the fixed guru setting under small w conditions. Besides, the imitation effect amplifies asymmetry and fat-tail characters of stock returns. We conclude that the communication network

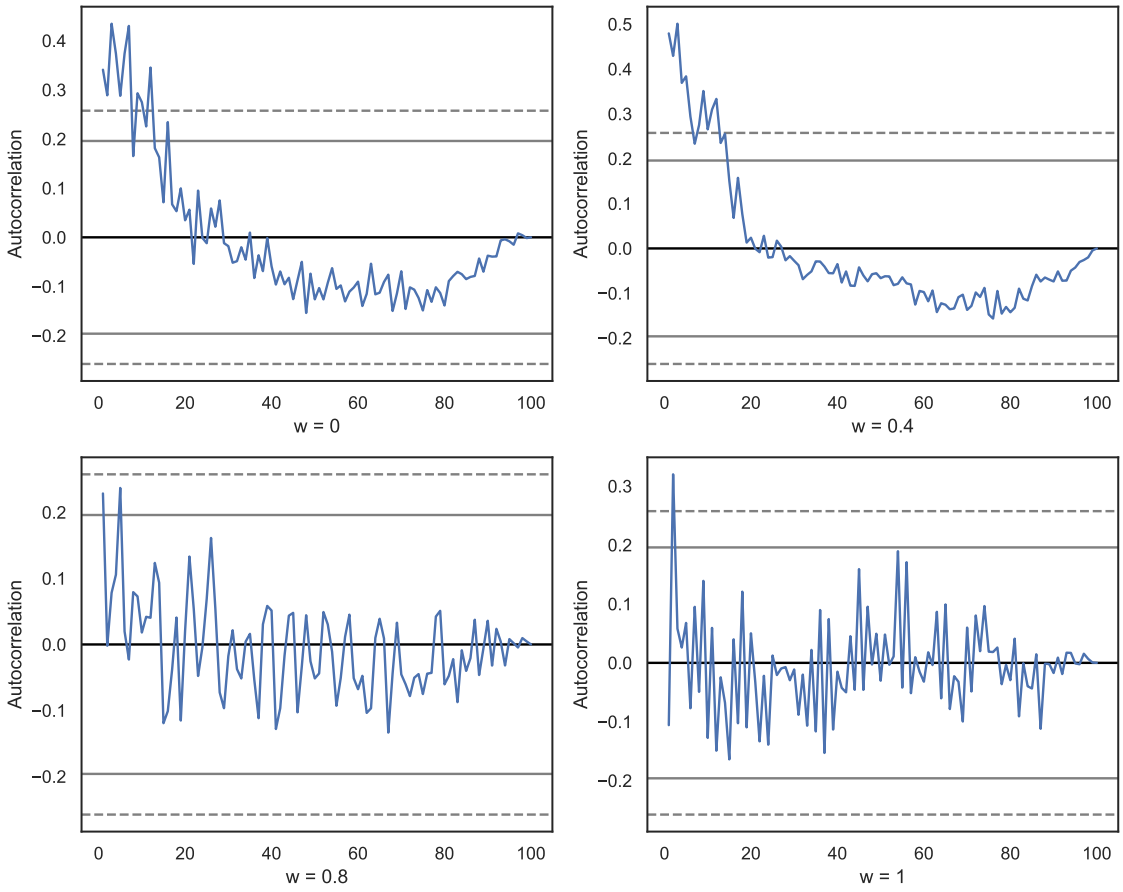


Figure 2.8: Auto-correlations of absolute stock returns under different w conditions.

is vital for our model to reproduce the stylised fact emerging in financial series.

2.7 Parameter Sensitivity Analysis

Due to the complexity of many ABMs, understanding the model dynamics is not a simple task. The sensitivity analysis can provide clues about how the parameters changes affect the model outcomes, which can help gain insight into how patterns and emergent properties are generated in ABMs. This section will perform a sensitivity analysis on our model, providing important implications for the universality of the simulation results. This analysis is prior to the calibration and validation process as a descriptive or in-sample exercise identifying the vital parameter for determining the statistical properties of the model.

At the very beginning, we investigate how the parameter w, A in Equation (2.7) and the number of agent N^{agent} affect the bid-ask spread of the stock price. We are

interested in a set of statistics $S = \{m_1, m_2, m_3, m_4\}$, which are the first four moments of simulated bid-ask spread $\{d_t, t = 1, \ldots, T\}$. For each run ($h = 1, 2, \ldots, H$), the simulation will produce a value for $m_i, i = 1, 2, 3, 4$. Due to the stochastic nature of the simulation, the value of m_i will be different for each independent run. Therefore, a sample set $\{m_i^1, m_i^2, \ldots, m_i^H\}$ for m_i containing H observations is expected by implementing H independent runs with the same parameter input. Then the distribution of $m_i, i = 1, 2, 3, 4$ can be roughly described by computing, for instance the mean $E(m_i)$ and the variance $V(m_i)$.

The parameter sensitivity analysis is based on the one-factor-at-a-time (OFAT) method, which essentially varies one parameter at a time and keeps all other parameters fixed. The use of OFAT can reveal the form of the relationship between the target parameter and the output. Especially, it can indicate whether the output is linear or nonlinear against the parameter. On the other hand, it can identify whether the tipping points exist, where the simulated output responds dramatically to a tiny parameter change.

For each run of the simulation, there are $T = 200$ periods and each period consists of $L = 300$ intra-day periods. The initial stock price is set as $P_0 = 100$. Each agent is assumed to have the same portfolio, which includes $C_0 = 10\,000$ cash and $S_0 = 100$ share. The parameter σ_0 is set as 1, which is used to generate the agent-specific volatility in Equation (2.7). Moreover, we run $H = 50$ independent simulations to obtain the distribution of a set of statistics S of the bid-ask spread.

2.7.1 Bid-ask Spread against Parameter w

We firstly look at how the parameter w affects the bid-ask spread of the simulated stock price. The value of w is evenly chosen starting from 0 to 1 with a step size $\Delta = 0.1$. The number of agent N^{agent} is set as 100 and the parameter A in Equation 2.7 equals to 0.2. The other parameters remain the same as stated in Section 2.7.

Figure 2.9 shows how the set of statistics $S = \{m_1, m_2, m_3, m_4\}$ of the bid-ask spread changes against parameter w. Each box and whisker plot contains $H = 50$ observations generated independently with the same parameter input, where the whisker is extended no more than $1.5 * \text{IQR}$ (the interquartile range = 75th percentile - 25th percentile) from the edges of the box, ending at the farthest data point within that

interval and outliers are plotted as separate dots.

The top-left panel illustrates that the sample value of the first moment of bid-ask spread $\{m_1^1, m_1^2, \ldots, m_1^H\}$ gradually decreases as the value of w increases, reaching the minimum around $w = 0.6$, then slightly climbing up. It is worth noting that there is no monotonic relationship between w and the mean of the bid-ask spread, and a hybrid structure of the agent expectation based on neighbour's and his own opinion makes the market more efficient with a low spread. Moreover, the variation of the bid-ask spread mean narrows as the value of w increases and the outliers vanish when $w \geq 0.8$. Similarly, the top-right panel displays the sample value of bid-ask spread standard deviation $\{m_2^1, m_2^2, \ldots, m_2^H\}$, which also decreases as the value of w increases. Meanwhile, the deviation of the outliers gets tight in the same way and disappears at $w = 1$. On the bottom side, the sample value of bid-ask spread skewness and kurtosis do not have an apparent decrease until $w \geq 0.5$ and the outliers keep existing for both of them.

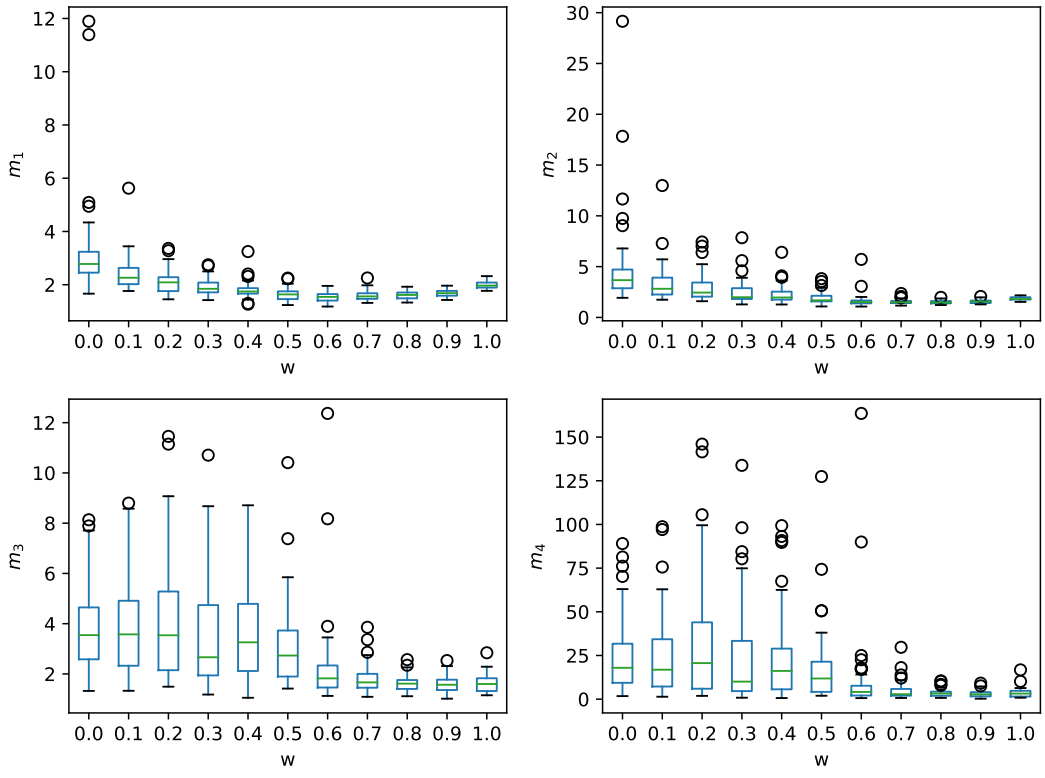


Figure 2.9: The box plots of the first four moments of bid-ask spread over the parameter w.

The box plot presented in Figure 2.9 implies that the imitation behaviour in the

market has a significant influence on the bid-ask spread. When agents heavily rely on others' opinions to generate their expectations, a one-side market is more likely to appear accompanied by a high bid-ask spread. By contrast, when a hybrid structure of the agent expectation dominates the model, the market becomes more stable and has higher liquidity with a low bid-ask spread.

Figure 2.10 illustrates how the change of parameter w affects the average performance and the variability of the first four moments of the bid-ask spread. The top-left panel shows a nonlinear relationship between the parameter w and the mean of the first moment of the bid-ask spread. As the value of w increases, the value of $E(m_1)$ steadily drops until reaching a minimum value around $w = 0.6$, then slowing climbing up. While the standard deviation of the first moment of bid-ask spread $SD(m_1)$ exhibits a dramatic decline when $w \leq 0.3$, which indicates that the tipping points exist for parameter w within the range. Then the value of $SD(m_1)$ presents a slow decay as $w \geq 0.3$, resulting from the fact that less probability of herding behaviour emerges in a simulation to produce irrational price dynamics. Likewise, the mean of the standard deviation of the bid-ask spread $E(m_2)$ at the top-right panel has a similar pattern to $E(m_1)$ showing a relative faster decline as the value of w increases, entering the valley around $w = 0.7$ and bounce back when $w = 1$. On the bottom side, there are obvious tipping points for both the mean of the skewness $E(m_3)$ and kurtosis $E(m_4)$ of the bid-ask spread in the range from $w = 0.4$ to $w = 0.8$. While their corresponding standard deviation $SD(m_3)$ and $SD(m_4)$ are more sensitive to the value of parameter w in the range from $w = 0.6$ to $w = 0.8$. This coordinates with the fact observed on the bottom side in Figure 2.9 that the deviation of the outliers rapidly decreases as $w \geq 0.6$.

The sensitivity analysis of parameter w based on the OFAT method indicates that a hybrid structure of agent expectation can create more market liquidity with a low bid-ask spread. A nonlinear relationship exists between the parameter w and the bid-ask spread. The variability of the simulation output is significantly affected by the value of w. Especially when w is close to 0, the herding effect is more likely to appear in the simulation due to the agent's expectation mainly accounts for his imitation behaviour. Therefore more outliers exist in the set of statistics $S = \{m_1, m_2, m_3, m_4\}$ of the bid-ask spread. Moreover, A more accurate sensitivity analysis can be implemented with a

smaller step size of parameter w, and a large number of independent runs can mitigate the stochastic effects.

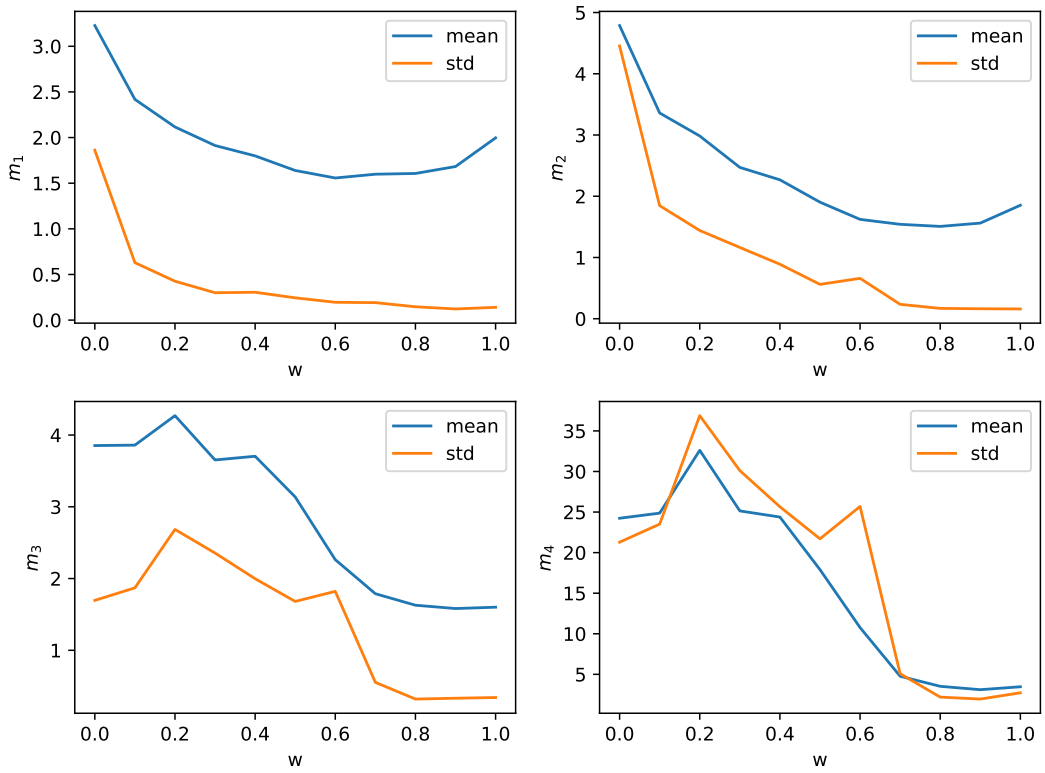


Figure 2.10: The sensitivity test of bid-ask spread statistics $S = \{m_1, m_2, m_3, m_4\}$ against the value of parameter w. The sample sets $\{m_i^h\}_{h=1}^H, i = 1, 2, 3, 4$ containing $H = 50$ independently generated observations are used to compute the corresponding mean $E(m_i)$ and standard deviation $SD(m_i)$.

2.7.2 Bid-ask Spread against Parameter A

After looking at how the parameter w affects the bid-ask spread in ABM simulations, we move on to implement a sensitivity analysis on parameter A, which directly scales the agent's specific volatility in Equation (2.7). The value of A is equally spaced in the range from 0.1 to 1 with a step size $\Delta = 0.1$. Due to the OFAT method, the value of w is set as 0.5, and the number of agents N^{agent} equals 100. The other parameters remain the same as stated in Section 2.7.

Figure 2.11 shows how the set of statistics $S = \{m_1, m_2, m_3, m_4\}$ of the bid-ask spread changes against parameter A. The top-left panel shows that there is a positive linear relationship between the median of each sample set $\{m_1^1, m_1^2, \ldots, m_1^H\}$ and the

parameter A, and the corresponding variation and outliers increase as the parameter A increases. A similar pattern can be observed for the sample set of m_2 with a low slope. By contrast, the sample set of m_3 and m_4 does not exhibit an obvious linear relationship. Moreover, the deviation of the outliers shrinks as A increases.

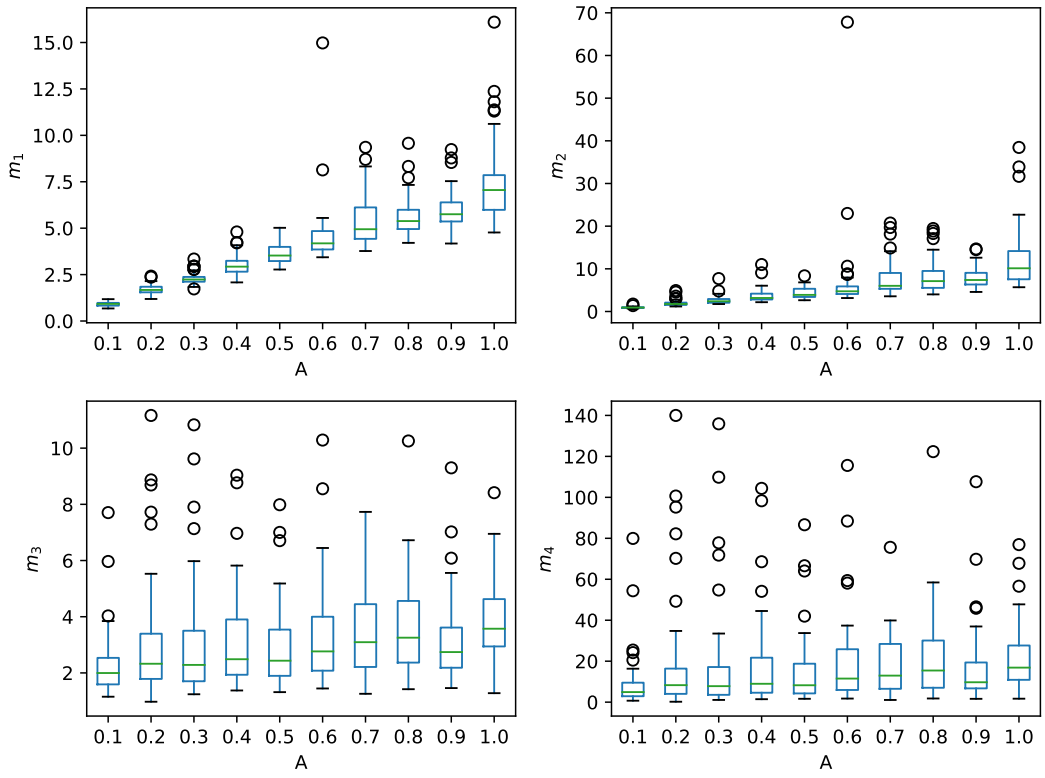


Figure 2.11: The box plots of the first four moments of bid-ask spread over the parameter A.

The box plot shown in Figure 2.11 indicates that high volatility of the agent's specific expectation on asset return will cause a high bid-ask spread in the ABM simulation. We can assert a linear relationship between the first two moments, m_1 and m_2, of the bid-ask spread and the parameter A. However, a quantified analysis is still required to support our findings.

Figure 2.12 displays how the change of parameter A affects the average performance and the variability of the first four moments of the bid-ask spread. The mean of the first two moments of the bid-ask spread, $E(m_1)$ and $E(m_2)$, show a linear relationship with the parameter A when $A \leq 0.6$. Then the slope of the corresponding curves become fluctuated. The tipping points exist for both the mean and standard deviation

of the third and fourth moments, $E(m_3)$ $E(m_4)$, $SD(m_3)$ and $SD(m_4)$, of the bid-ask spread when $A \leq 0.2$. Later, they all show an oscillating movement due to the stochastic nature of the agent-based simulation. The standard deviation of the first two moments of the bid-ask spread, $SD(m_1)$ and $SD(m_2)$, can only show a positive relationship with the parameter A when $A \leq 0.4$.

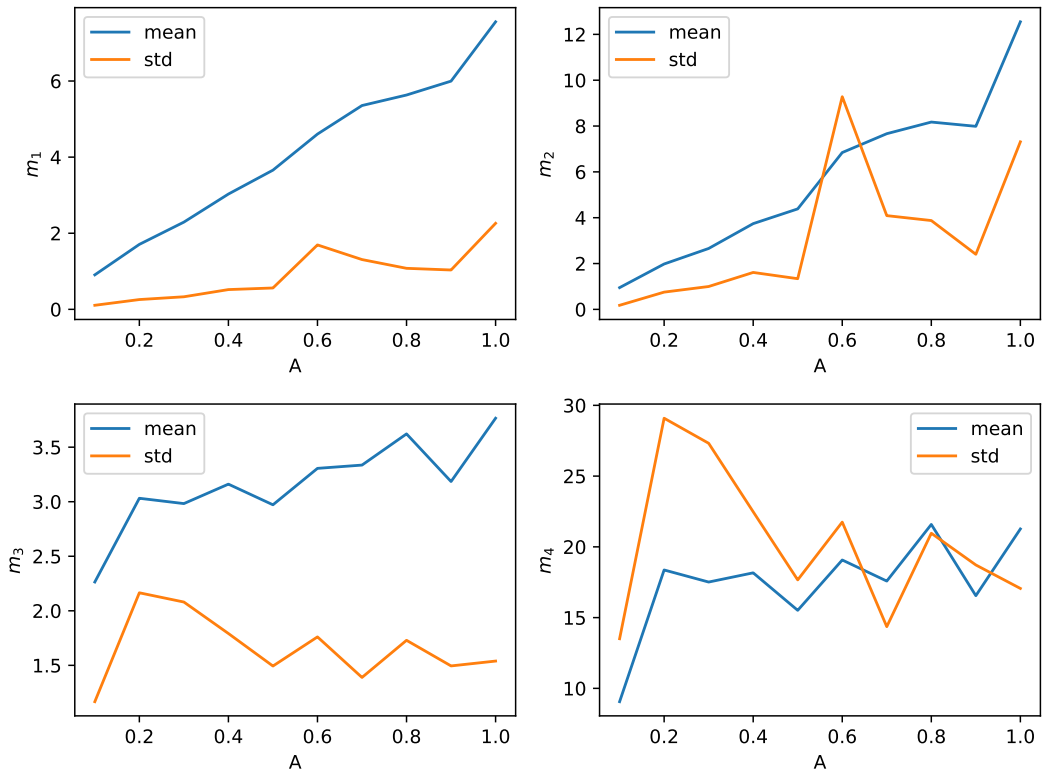


Figure 2.12: The sensitivity test of bid-ask spread statistics $S = \{m_1, m_2, m_3, m_4\}$ against the value of parameter A. The sample sets $\{m_i^h\}_{h=1}^H, i = 1, 2, 3, 4$ containing $H = 50$ independently generated observations are used to compute the corresponding mean $E(m_i)$ and standard deviation $SD(m_i)$.

The sensitivity analysis of how the bid-ask spread responds to the change of parameter A and it suggests a better control of the value of parameter A within the range of $(0, 0.4]$, where a linear property can be obtained for the mean and standard deviation of the first two moments of the bid-ask spread. Furthermore, to gain insight into the properties of the standard deviation of the corresponding moments, a smaller step size of parameter A is required to get a more accurate analysis.

2.7.3 Bid-ask Spread against the Number of Agents

At last, we investigate the effect of the number of agents N^{agent} on the bid-ask spread in ABM simulations. We uniformly choose the value of N^{agent} in the range from 100 to 1 000 with a step size $\Delta = 100$. Besides, the value of parameters w and A is set as 0.5 and 0.2, respectively. The other parameters remain the same as stated in Section 2.7.

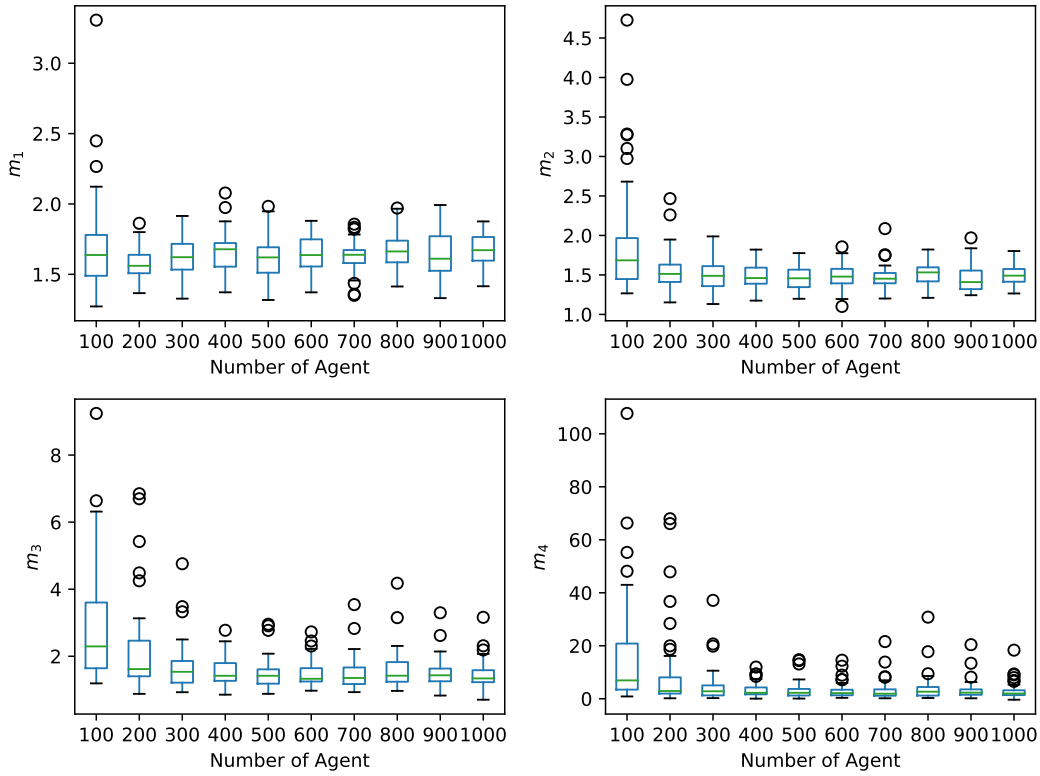


Figure 2.13: The box plots of the first four moments of bid-ask spread over the number of agents N^{agent}.

Figure 2.13 shows how the set of statistics $S = \{m_1, m_2, m_3, m_4\}$ of the bid-ask spread changes against the number of agents N^{agent}, which do not have an apparent influence on both the first and the second moment, m_1 and m_2, of the bid-ask spread. By contrast, although many outliers exist for the distribution of m_3 and m_4, the corresponding median and variance decay slowly as the number of agents N^{agent} increases.

Furthermore, Figure 2.14 exhibits how the change of the number of agents N^{agent} influences the average performance and the variability of the first four moments of the bid-ask spread. Expect that there is a sudden drop for $E(m_1)$, $E(m_2)$, $SD(m_1)$

and $SD(m_2)$ when the number of agents N^{agent} increases from 100 to 200, the bid-ask spread is not affected as more agents are included in the model. While a similar pattern can be observed for $E(m_3)$, $E(m_4)$, $SD(m_3)$ and $SD(m_4)$ when the number of agents $N^{agent} \leq 400$.

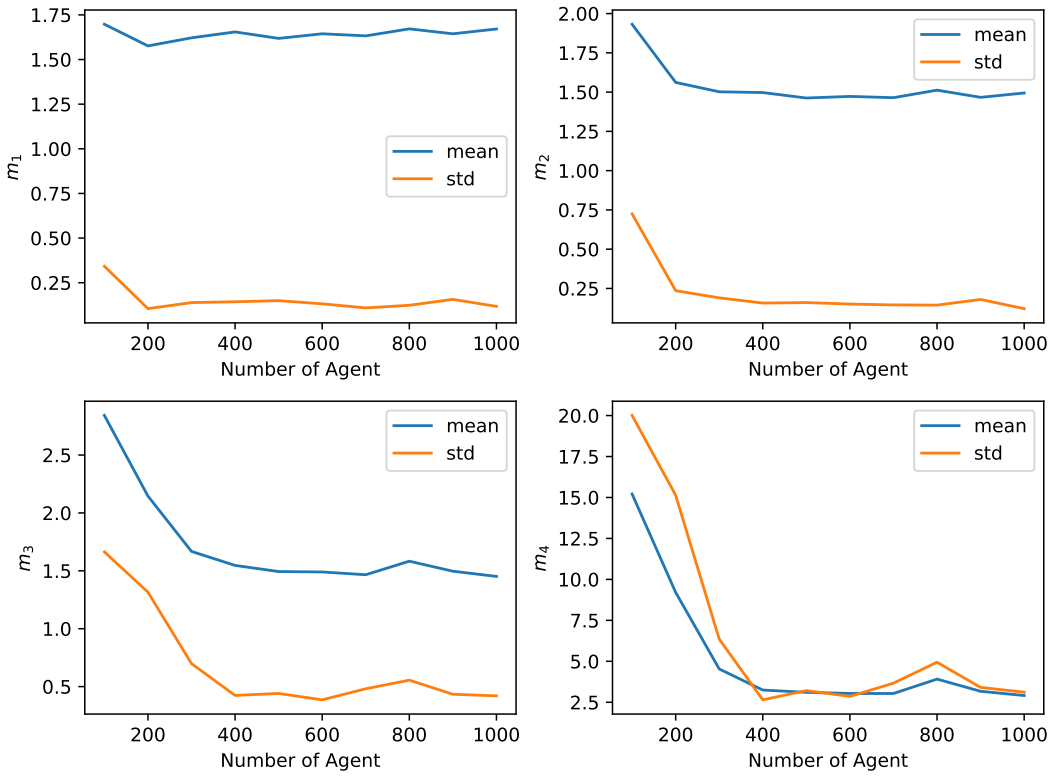


Figure 2.14: The sensitivity test of bid-ask spread statistics $S = \{m_1, m_2, m_3, m_4\}$ against the number of agents N^{agent}. The sample sets $\{m_i^h\}_{h=1}^{H}, i = 1, 2, 3, 4$ containing $H = 50$ independently generated observations are used to compute the corresponding mean $E(m_i)$ and standard deviation $SD(m_i)$.

The sensitivity analysis of how the bid-ask spread responds to the number of agents N^{agent} implies that the bid-ask spread can be narrowed if we introduce new agents to the market consisting of a small size of participants. However, when there are sufficient candidates in the market, the set of statistics $S = \{m_1, m_2, m_3, m_4\}$ of the bid-ask spread will not be influenced by the number of agents N^{agent} anymore.

2.8 Summary

In this chapter, we followed Tedeschi's [52] (2009) framework to construct an agent-based model with noisy agents in the stock market. By involving an endogenous

mechanism of imitation, the herding effect can be generated that wealthy agents are more likely to be imitated by others, causing a larger fluctuation in price. Due to the model being a closed system, the price movement will be restricted to oscillate around a constant P^{mean}, which is the ratio of the total money supply to total share supply in the system. Meanwhile, the simulated log-return is verified to exhibit a leptokurtic, asymmetric and fat-tailed distribution. Other stylised empirical facts emerging in real financial prices series, such as zero auto-correlation of log returns and volatility clustering, exist in the model outputs.

In this thesis, we contributed to implementing a parameter sensitivity analysis based on the one-factor-at-a-time method, which is prior to the calibration process in Chapter 4 as a descriptive or in-sample exercise identifying the vital parameter for determining the statistical properties of the model. Based on the analysis, we found that a low bid-ask spread of simulated price can be obtained with a hybrid structure of agent expectation, where more market liquidity can be created. The herding effect is more likely to be observed when the agent's expectation mainly accounts for his imitation behaviour. Within a suitable range, the scale parameter A of the agent's specific volatility shows a linear relationship against the first and second moment of the bid-ask spread. Moreover, additional participants will not significantly influence the bid-ask spread when the market has sufficient agents.

Chapter 3

Stock Market with Trend Following Agents

In Chapter 2, we have introduced a basic agent-based model, which can reproduce a set of stylised empirical facts only with some necessary mechanisms, including the simplest noise agent assumption. In this section, we want to equip agents with an intelligent strategy and investigate market performance under the condition of these more intelligent agents.

3.1 Portfolio Choice

Markowitz [61] (1952) firstly proposed the portfolio choice problem that an agent wants to create a portfolio with a maximum probability of expected return under a fixed level of variance. Although the original model is built in a mean-variance framework, it can also be written within the Von Neumann-Morgenstern expected utility framework. The generalised Markowitz problem is extended to multi-period versions by Samuelson [62] (1969) with a discrete approach and Merton (1969) with a continuous approach (see [63] and [64]). Then the portfolio choice problem can be treated as an optimal investment or consumption problem.

Many classic Merton problems consider the input parameters as known constants or stochastic processes with known distributions, which ignores the uncertainty for the price return to follow a given distribution. Therefore it is a valuable extension by involving the learning technique in the optimal allocation model. In [65], Karatzas

and Zhao (2001) made an outstanding contribution by using martingale methods under filtration to solve the optimal allocation problem for almost any utility function with an assumption that stock return follows a Gaussian distribution with unknown mean.

In this chapter, we first review the work of Bismuth *et al.* [53] (2019), in which a continuous portfolio choice problem is considered under the assumption that the expected return of risky assets is unknown. They used conjugate priors and updated the corresponding parameters with a simple Markovian approach. A closed-form solution of the Hamilton-Jacobi-Bellman(HJB) equation can be obtained in a stochastic optimal control problem with constant relative risk aversion (CRRA) and constant absolute risk aversion (CARA) utility function. Then, an agent-based model with trend-following agents is established with some necessary notations. Next, we present the order placement mechanism in the CARA and CRRA cases. An illustration of how the Bayesian learning technique follows the asset return is also provided. The simulation results are displayed under the case that agents have a CRRA utility function with $\gamma = 1$, and we contribute to illustrating how a market can be set up with a deterministic trend and what leads to a crash in the system. Finally, the parameter sensitivity analysis is presented to find the suitable initial condition and constraint of parameters for a traceable and reproducible simulation to exist.

3.2 Stock Price Dynamics

We follow the model introduced by Bismuth *et al.* [53] (2019), considering a portfolio allocating problem faced by an agent in a simplified financial market only consisting of one risk-free asset and one risky asset. Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t^W)_{t\in R_+}, \mathbb{P}\right)$ be a filtered probability space with $\mathcal{F}_t^W$ generated by the Wiener process W_t.

The risky free interest rate is donated by r and the risky asset P has classical log-normal dynamics

$$dP_t = \mu P_t dt + \sigma P_t dW_t, \tag{3.1}$$

where the drift μ is unknown and $\sigma > 0$.

3.3 Bayesian Learning

At the time $t = 0$, agents' belief about the value of μ is assumed as a Gaussian prior distribution

$$\mu \sim \mathbf{N}(\beta_0, v_0^2), \tag{3.2}$$

where $\beta_0 \in R$ and $v_0^2 > 0$.

A classical result of filtering methods shows that the conditional distribution of μ given $\mathcal{F}_t^P$ is Gaussian and $(\mathcal{F}_t^P)_{t\in R_+}$ is denoted as a filtration generated by $(P_t)_{t\in R_+}$. Specifically, we have that

Proposition 1. *Let $t \in R_+$, given $\mathcal{F}_t^P$, μ has a conditional distribution with mean β_t and variance v_t^2, where*

$$\beta_t = \frac{\sigma^2}{\sigma^2 + v_0^2 t}\beta_0 + \frac{v_0^2}{\sigma^2 + v_0^2 t}(\mu_t + \sigma W_t), \tag{3.3}$$

and

$$v_t^2 = \frac{\sigma^2 v_0^2}{\sigma^2 + v_0^2 t}. \tag{3.4}$$

The proof of Proposition 1 is well presented in Appendix A, and we notice that $\mu_t + \sigma W_t = \log\left(\frac{P_t}{P_0} + \frac{1}{2}\sigma^2 t\right)$, which is observable at time t.

In Proposition 1, the process $(v_t)_t$ is a deterministic process and the process $(\beta_t)_t$ is a stochastic process having the following dynamics

$$\begin{aligned} d\beta_t &= g(t)(\mu - \beta_t)dt + \sigma g(t) dW_t \\ &= \sigma g(t) d\hat{W}_t, \end{aligned} \tag{3.5}$$

where the function g is defined by

$$g : t \in R_+ \mapsto \frac{v_0^2}{\sigma^2 + v_0^2 t}, \tag{3.6}$$

and the process $\left(\hat{W}_t\right)_t$ is defined by

$$\hat{W}_t = W_t + \int_0^t \frac{\mu - \beta_s}{\sigma} ds, \tag{3.7}$$

which is a Wiener process adapted to $\left(\mathcal{F}_t^P\right)_{t\in R_+}$ (see proof in Appendix A).

3.4 Optimal Portfolio Choice in CARA Case

The strategy of the agent is implemented by a process $(M_t)_t$ representing the amount invested in the risky asset, so $M_t = S_t P_t$. Therefore the value of agent's portfolio is modelled by a process $(V_t)_t$ with $V_0 > 0$ and the dynamic

$$\begin{aligned} dV_t &= ((\mu - r)M_t + rV_t)dt + \sigma M_t dW_t \\ &= ((\beta_t - r)M_t + rV_t)dt + \sigma M_t d\hat{W}_t. \end{aligned} \tag{3.8}$$

We define the set of admissible strategies as $\mathcal{A} = \mathcal{A}_0$ where

$$\mathcal{A}_t = \{(M_s)_{s\in[t,T]}, M \text{ is measurable}, \mathcal{F}^P\text{-adapted under the linear growth condition}\} \tag{3.9}$$

A measurable and $\mathcal{F}^P$-adapted process $(V_t)_t$ is said to satisfy a linear growth with respect to $\hat{W}$, if for all $t \in [0, T]$

$$|V_t| \leq C_T \left(1 + \sup_{s\in[0,t]} |\hat{W}_s|\right), \tag{3.10}$$

where C_T is deterministic and depends only on T.

For $t \in R_+$, given $M \in \mathcal{A}_t$ and $s \geq t$, we define that

$$\beta_s^{t,\beta} = \beta + \int_t^s \sigma g(t^{'}) d\hat{W}t^{'} \tag{3.11}$$

$$V_s^{t,V,\beta,M} = V + \int_t^s ((\beta_{t'}^{t,\beta} - r)M_{t'} + rV_{t'}^{t,V,\beta,M})dt^{'} + \int_t^s \sigma M_{t'} d\hat{W}t^{'}. \tag{3.12}$$

The optimal portfolio choice problem in CARA case origins from an agent who wants to maximise their utility function

$$\sup_{M\in\mathcal{A}_t} E^Q[-\exp(-\gamma V_T^{0,V_0,\beta_0,M})], \tag{3.13}$$

and the value function u associated with this problem is defined by

$$u(t, V, \beta) = \sup_{M\in\mathcal{A}_t} E^Q[-\exp(-\gamma V_T^{t,V,\beta,M})]. \tag{3.14}$$

Then the HJB for this problem is written as

$$\begin{aligned} -\partial_t u - \frac{1}{2}\sigma^2 g(t)^2 \partial^2_{\beta\beta} u - \sup_{M\in\mathcal{A}_t} \{((\beta - r)M_t + rV)\partial_V u + \\ \frac{1}{2}\sigma^2 M^2 \partial^2_{VV} u + \sigma^2 M g(t) \partial^2_{\beta V} u\} = 0, \end{aligned} \tag{3.15}$$

with the terminal condition

$$u(T, V, \beta) = -\exp(-\gamma V). \tag{3.16}$$

3.4.1 Solve the HJB Equation in CARA Case

To solve the HJB equation (3.15) with terminal condition (3.16), the following ansatz is used

$$u(t, V, \beta) = -\exp\left[-\gamma\left(e^{r(T-t)}V + \varphi(t, \beta)\right)\right]. \tag{3.17}$$

Proposition 2. *Suppose there exists* $\varphi \in \mathcal{C}^{1,2}\left([0, T] \times R\right)$ *satisfying*

$$-\partial_t\varphi(t, \beta) - \frac{1}{2}\sigma^2 g^2(t)\partial^2_{\beta\beta}\varphi(t, \beta) - \frac{(\beta - r)^2}{2\gamma\sigma^2} + g(t)(\beta - r)\partial_\beta\varphi(t, \beta) = 0, \tag{3.18}$$

with the terminal condition

$$\varphi(T, \beta) = 0. \tag{3.19}$$

Then the u *defined in Equation (3.17) is solution of the HJB equation (3.18) with terminal condition (3.19), and the supremum of Equation (3.15) can be obtained at*

$$M^* = e^{-r(T-t)}\left(\frac{\beta - r}{\gamma\sigma^2} - g(t)\partial_\beta\varphi(t, \beta)\right). \tag{3.20}$$

The Proposition 2 has converted the three-variable nonlinear PDE problem into a two-variable linear one. The next step boils down the (3.18) with terminal condition (3.19) to a system of first-order linear ODEs.

Proposition 3. *Suppose that* $(a, b) \in (\mathcal{C}^1([0, T]))^2$ *satisfies the following system of ODEs:*

$$\begin{cases} a^{'}(t) + \frac{1}{2}\sigma^2 g^2(t)b(t) & = 0 \\ b^{'}(t) + \frac{1}{\gamma\sigma^2} - 2g(t)b(t) & = 0, \end{cases} \tag{3.21}$$

with terminal conditions

$$\begin{cases} a(T) & = 0 \\ b(T) & = 0. \end{cases} \tag{3.22}$$

Then the φ *defined by*

$$\varphi(t, \beta) = a(t) + \frac{1}{2}b(t)(\beta - r)^2, \tag{3.23}$$

satisfies (3.18) with terminal condition (3.19).

The system of ODEs (3.21) with terminal conditions (3.22) can be solved easily in the following proposition.

Proposition 4. *The function $a(t)$ and $b(t)$ defined by*

$$\begin{cases} a(t) &= \frac{1}{2\gamma}\left(\log\left(\frac{g(t)}{g(T)}\right) - (T-t)g(T)\right) \\ b(t) &= \frac{1}{\gamma\sigma^2}(T-t)\frac{g(T)}{g(t)}, \end{cases} \tag{3.24}$$

satisfies the system (3.21) with terminal conditions (3.22).

Combining the notations stated in Propositions 2, 3 and 4, the function u defined by

$$\begin{aligned} & u(t,V,\beta) \\ =& -\exp\left[-\gamma\left(e^{r(T-t)}V + \varphi(t,\beta)\right)\right] \\ =& -\exp\left[-\gamma\left(e^{r(T-t)}V + a(t) + \frac{1}{2}b(t)(\beta-r)^2\right)\right] \\ =& -\exp\left[-\gamma\left(e^{r(T-t)}V + \frac{1}{2\gamma}\left(\log\left(\frac{g(t)}{g(T)}\right) - (T-t)g(T)\right) + \right.\right. \\ & \left.\left.\frac{1}{2\gamma\sigma^2}(T-t)\frac{g(T)}{g(t)}(\beta-r)^2\right)\right], \end{aligned} \tag{3.25}$$

satisfies the HJB equation (3.15) with terminal condition (3.16), and the supremum of $u(t,V,\beta)$ is obtained at

$$M^* = e^{-r(T-t)}\frac{g(T)}{g(t)}\frac{\beta-r}{\gamma\sigma^2}. \tag{3.26}$$

The argument of the maximum in (3.15) is obtained by plugging the expression of φ in (3.20).

In particular, agents equipped with the optimal portfolio choice strategy with a CARA utility function should invest the amount of M^* cash in risky asset at time t. The proofs of Proposition 2 and 3 can be found in Appendix A. While the proof of Proposition 4 is straightforward by differentiating (3.24) and verifying the terminal conditions.

3.5 Optimal Portfolio Choice in CRRA Case

In the CRRA case, the agent's strategy is modelled by the process $(\theta_t)_t$, which describes the proportion of the portfolio value invested in the risky asset. Then the dynamic of

the agent's portfolio value is

$$\begin{aligned} dV_t &= ((\mu - r)\theta_t V_t + rV_t)dt + \sigma\theta_t V_t dW_t \\ &= ((\beta - r)\theta_t V_t + rV_t)dt + \sigma\theta_t V_t d\hat{W}_t. \end{aligned} \tag{3.27}$$

Let γ denote the relative risk aversion parameter of agents. The set of admissible strategies depends on the value of γ.

For all $t \in [0, T]$, if $\gamma < 1$, the set of strategies is introduced as

$$\mathcal{A}_t^\gamma = \left\{ (\theta_s)_{s\in[t,T]}, \theta \text{ is measurable and } \mathcal{F}^P\text{-adapted, } \mathbb{P}\left(\int_t^T \theta_s^2 ds < \infty = 1 \right) \right\}.$$

If $\gamma \geq 1$, the set of strategies is denoted as

$$\begin{aligned} \mathcal{A}_t^\gamma = \Big\{ &(\theta_s)_{s\in[t,T]}, \text{measurable, } \mathcal{F}^P\text{-adapted, and satisfying} \\ &\text{the linear growth condition with respect to } \hat{W} \Big\}. \end{aligned}$$

Therefore the set of admissible strategies is $\mathcal{A}^\gamma = \mathcal{A}_0^\gamma$. For $t \in R_+$, given $\theta \in \mathcal{A}_t^\gamma$ and $s \geq t$, we define

$$\beta_s^{t,\beta} = \beta + \int_t^s \sigma g(t')d\hat{W}_{t'} \tag{3.28}$$

$$V_s^{t,V,\beta,\theta} = V + \int_t^s ((\beta_{t'}^{t,\beta} - r)\theta_{t'} + r)V_{t'}^{t,V,\beta,\theta}dt' + \int_t^s \sigma\theta_{t'}V_{t'}^{t,V,\beta,\theta}d\hat{W}_{t'}. \tag{3.29}$$

The optimal strategy for the agent in CRRA case is to maximise the utility function

$$\sup_{\theta\in\mathcal{A}^\gamma} E^Q[U^\gamma(V_T^{0,V_0,\beta_0,\theta})] \tag{3.30}$$

where the CRRA utility function is defined by

$$U^\gamma(V) = \begin{cases} \frac{V^{1-\gamma}}{1-\gamma} & \text{if } \gamma \neq 1 \\ \log(V) & \text{if } \gamma = 1. \end{cases} \tag{3.31}$$

The value function u associate with this problem is defined by

$$u(t, V, \beta) = \sup_{\theta\in\mathcal{A}_t^\gamma} E^Q[U^\gamma(V_T^{t,V,\beta,\theta})] \tag{3.32}$$

Then HJB for this problem is then written as

$$\begin{aligned} -\partial_t u(t,V,\beta) - \frac{1}{2}\sigma^2 g^2(t)\partial_{\beta\beta}^2 u(t,V,\beta) - \sup_\theta \{((\beta - r)\theta + r)V\partial_V u(t,V,\beta) \\ + \frac{1}{2}\sigma^2\theta^2 V^2\partial_{VV}^2 u(t,V,\beta) + \sigma^2 g(t)\theta V\partial_{V\beta}^2 u(t,V,\beta)\} = 0 \end{aligned} \tag{3.33}$$

with the terminal condition

$$u(T, V, \beta) = U^\gamma(V) \tag{3.34}$$

3.5.1 Solve the HJB Equation in CRRA Case

For $\gamma \neq 1$, the following ansatz is used to solve HJB equation (3.33) with terminal condition (3.34)

$$u(t, V, \beta) = \frac{\left(e^{r(T-t)}V\right)^{1-\gamma}}{1-\gamma} \exp[\varphi(t, \beta)]. \tag{3.35}$$

Proposition 5. *For $\gamma \neq 1$, assume there exists $t' \in [0, T)$ and $\varphi \in \mathcal{C}^{1,2}([t', T] \times R)$ satisfying*

$$\begin{aligned} -\frac{1}{1-\gamma}\partial_t\varphi(t,\beta) - \frac{1}{2(1-\gamma)}\sigma^2 g^2(t)\partial^2_{\beta\beta}\varphi(t,\beta) - \frac{1}{2\gamma(1-\gamma)}\sigma^2 g^2(t)(\partial_\beta\varphi(t,\beta))^2 \\ -\frac{1}{\gamma}\frac{(\beta-r)^2}{2\sigma^2} - \frac{1}{\gamma}g(t)(\beta-r)\partial_\beta\varphi(t,\beta) = 0, \end{aligned} \tag{3.36}$$

with terminal condition

$$\varphi(T, \beta) = 0. \tag{3.37}$$

Then function u defined in (3.35) is the solution of HJB equation (3.33) with terminal condition (3.34) on $[t', T] \times R^*_+ \times R$. Furthermore, the supremum in (3.33) can be obtained at

$$\theta^* = \frac{\beta - r}{\gamma\sigma^2} + \frac{1}{\gamma}g(t)\partial_\beta\varphi(t,\beta). \tag{3.38}$$

Similar to the solving procedure in the CARA case, the Proposition 5 has converted the three-variable HJB PDE into a two-variable PDE. In the following proposition, we will boil down the new PDE to a system of ODEs. It is worthy to note that both the two-variable PDE and the system of ODEs are nonlinear.

Proposition 6. *For $\gamma \neq 1$, suppose there exists $t' \in [0, T]$ and $(a, b) \in \left(\mathcal{C}^1([t', T])\right)^2$ satisfying the following system of ODEs*

$$\begin{cases} a'(t) + \frac{1}{2}\sigma^2 g^2(t) b(t) & = 0 \\ b'(t) + \frac{1}{\gamma}\sigma^2 g^2(t) b^2(t) + \frac{1-\gamma}{\gamma}\frac{1}{\sigma^2} + 2\frac{1-\gamma}{\gamma}g(t)b(t) & = 0, \end{cases} \tag{3.39}$$

with terminal conditions

$$\begin{cases} a(T) & = 0 \\ b(T) & = 0. \end{cases} \tag{3.40}$$

Then the function φ defined by

$$\varphi(t, \beta) = a(t) + \frac{1}{2}b(t)(\beta - r)^2, \tag{3.41}$$

satisfies (3.36) with terminal condition (3.37).

Note that the second ODE in (3.39) is a Riccati equation, which has a particular solution $-\frac{1}{\sigma^2 g}$, and a classic solution of b is to assume it has a form

$$b(t) = -\frac{1}{\sigma^2 g(t)} + d(t). \tag{3.42}$$

Then Equation (3.42) is the solution of ODE in (3.39) with terminal condition $b(T) = 0$ if and only if d solves the following equation

$$d^{'}(t) + \frac{1}{\gamma}\sigma^2 g^2(t) d^2(t) - 2g(t)d(t) = 0 \tag{3.43}$$

with terminal condition $d(T) = \frac{1}{\sigma^2 g}$.

By using a simple transform $e = \frac{1}{d}$, Equation (3.43) can be reduced to a first-order linear ODE

$$e^{'}(t) + 2g(t)e(t) = \frac{1}{\gamma}\sigma^2 g^2(t), \tag{3.44}$$

with terminal condition $e(T) = \sigma^2 g(T)$, and it has a unique solution

$$e(t) = \frac{1}{\gamma}\sigma^2 g(t) + \left(1 - \frac{1}{\gamma}\right)\sigma^2 \frac{g^2(t)}{g(T)}. \tag{3.45}$$

Proposition 7. *For $\gamma \neq 1$, we define that*

$$\tilde{t} = (1-\gamma)T - \gamma\frac{\sigma^2}{v_0^2}, \tag{3.46}$$

and

$$I = \begin{cases} [0, T] & \text{if } \tilde{t} < 0 \ (i.e., \gamma > 1, \text{ or } \gamma < 1 \text{ and } \frac{\sigma^2}{v_0^2} > \frac{1-\gamma}{\gamma}T), \\ (\tilde{t}, T] & \text{otherwise.} \end{cases} \tag{3.47}$$

Then the function a and b defined on I

$$\begin{cases} a(t) &= \frac{\gamma}{2}\log\left(\frac{\gamma g(t)}{(\gamma-1)g(t)+g(T)}\right) + \frac{1}{2}\log\left(\frac{g(T)}{g(t)}\right) \\ b(t) &= \frac{1-\gamma}{\sigma^2}\frac{1}{g(t)}\frac{g(t)-g(T)}{(\gamma-1)g(t)+g(T)}. \end{cases} \tag{3.48}$$

satisfies the system of ODEs (3.39) with terminal condition(3.40).

Combining the Proposition 5, 6 and 7, for $\gamma \neq 1$, the function u defined on $I \times$

$R_+^* \times R$

$$\begin{aligned} u(t,V,\beta) =& \frac{\left(e^{r(T-t)}V\right)^{1-\gamma}}{1-\gamma}\exp[\varphi(t,\beta)] \\ =& \frac{\left(e^{r(T-t)}V\right)^{1-\gamma}}{1-\gamma}\exp[a(t)+\frac{1}{2}b(t)(\beta-r)^2] \\ =& \frac{(e^{r(T-t)}V)^{1-\gamma}}{1-\gamma}\exp[\frac{\gamma}{2}\log(\frac{\gamma}{(\gamma-1)+\frac{g(T)}{g(t)}})+\frac{1}{2}\log(\frac{g(T)}{g(t)}) \\ &+\frac{1}{2}\frac{(1-\gamma)}{\sigma^2}\frac{1}{g(t)}\frac{g(t)-g(T)}{(\gamma-1)g(t)+g(T)}(\beta-r)^2], \end{aligned} \tag{3.49}$$

satisfies the HJB equation (3.33) on $I \times R_+^* \times R$ with terminal condition (3.34) and the supremum in (3.33) can be obtained at

$$\theta^* = \frac{\beta-r}{\gamma\sigma^2}\frac{\gamma g(T)}{(\gamma-1)g(t)+g(T)}. \tag{3.50}$$

Proposition 8. *For $\gamma = 1$, the function u defined by*

$$u(t,V,\beta) = \log(e^{r(T-t)}V), \tag{3.51}$$

satisfies the equation (3.33) with terminal condition (3.34) and the supremum in (3.33) is obtained at

$$\theta^* = \frac{\beta-r}{\sigma^2}. \tag{3.52}$$

For the case $\gamma = 1$, the solution of the HJB equation (3.33) has no dependence on β. In particular, agents equipped with the optimal portfolio choice strategy with a CRRA utility function should invest a proportion of θ^* cash in risky asset at time t. The proofs of Proposition 5, 6 be found in Appendix A. While the proof of Proposition 7 is straightforward by differentiating (3.48) and verifying the terminal condition.

3.6 Notations for the Model

Although the optimal strategy introduced above is built on a continuous-time, our model will use a discrete-time with T periods for simplicity. A population of $N^{learning}$ learning agents are added into the market and randomly picked up to make their operations at each intra-day period.

Let t_k be denoted as the period, which starts from t_1 and end with t_T ($T < \infty$) and the initial period is denoted as t_0. Within each period, there are L intra-day periods

denoting as t_{k_l}, then t_{k_1} and t_{k_L} represents the first and the last intra-day period t_k respectively.

Assume that agent i has an account book, which consists of the following 6 items:

- *Cash.* C_t^i represents the total money agent i holding at time t, which will increase/decrease by the amount of cash in a completed buy/sell transaction.
- *Limit Cash.* It represents the amount of cash occupied by limit bid orders submitted by agent i. The limit cash will decrease when limit orders are executed or cancelled.
- *Available Cash.* A_t^i represents the amount of cash that agent i can use to buy shares. The available cash will increase when shares are sold, or the limit bid orders are cancelled. On the contrary, the available cash will decrease when shares are bought, or the limit bid orders are submitted.
- *Share.* S_t^i represents the number of shares held by agent i at time t, which will increase/decrease by the amount of share in a completed buy/sell transaction.
- *Limit Share.* It represents the share occupied by limit ask orders, which will decrease when limit orders are executed or cancelled.
- *Available Share.* N_t^i represents the number of shares that agent i can sell, which will increase when bid orders are completed, or limit ask orders are cancelled. While both market and limit ask orders will decrease the available share.

At the initial time, all agents are provided with an amount of cash C_0 and S_0 shares. Given that P_t is denoted as the asset price at time t, the value of agent i's portfolio is denoted as V_t^i with the initial value $V_0^i = C_0^i + S_0^i P_0$.

We assume that all agents have a common understanding of the dynamic of stock price, which is written as

$$dP_t = \mu P_t dt + \sigma P_t dW_t, \tag{3.53}$$

where $(W_t)_t$ is Wiener process.

Let β_t be denoted as the agent's belief about the value of μ at time t. Agents will use the Bayesian learning technique to update their latest knowledge of β_t by observing the current stock price P_t.

In the CARA case, the amount of agent's investment in risky asset is denoted as $M_t = S_t P_t$ at time t. While his optimal investment is denoted as M_t^*. Similarly, in the CRRA case, the proportion of agent's investment in a risky asset at time t is denoted as θ_t with corresponding optimal proportion written as θ_t^*.

The limit order book collects all orders submitted by agents. Since the order is submitted at each intra-day period, then the corresponding bid and ask order price is defined by $b_{t_{k_l}}$ and $a_{t_{k_l}}$ respectively. While the volume of the order is represented by $h_{t_{k_l}}$.

3.7 Place Orders

We have introduced the optimal strategies for solving portfolio choice problems using CARA and CRRA utility functions. This part will combine these strategies with our Agent-based model by modifying the expectation formation mechanism and order submission for each agent. We assume that agents can only buy or sell the share with their holding cash and inventories.

3.7.1 CARA Case

The order placement for agents with the CARA utility function is straightforward. At the beginning of the simulation, agent i generates his origin belief β_0 about the value of drift μ with a Gaussian distribution $\beta_0 \sim \mathbf{N}\left(\mu_{\beta_0}, \sigma_{\beta_0}^2\right)$. The agent i's optimal strategy with the CARA utility function is to hold a amount of money M^* (Equation (3.26)) invested in the risky asset. As no credit transaction is allowed in the system, each agent's optimal investment in the risky asset is bounded in $M_t^* \in [0, M_t + C_t]$.

Let P_{t_k} represent the opening price at period t_k. During the period, if $M_{t_{k_l}} < M_{t_k}^*$, where $M_{t_k}^*$ is the amount of money the agent want to invest in risky asset at the end of the period t_k, the agent will submit limit bid orders to increase their investment in the risky asset at the following intra-day periods. The limit bid order price $b_{t_{k_l}}$ is assumed to have a formula

$$b_{t_{k_l}} \sim U(P_{t_k}, P_{t_k} e^{(\beta_{t_k} - \frac{1}{2}\sigma^2)\Delta t + \sigma\xi\sqrt{\Delta t}}), \tag{3.54}$$

where $\xi \sim N(0,1)$ is a standard normal variable and the time step $\Delta t = 1$. In order

to get the optimal position of risky asset investment $M^*_{t_k}$, the limit bid order volume $h_{t_{k_l}}$ can be calculated by

$$h_{t_{k_l}} = \frac{M^*_{t_k} - M_{t_{k_l}}}{b_{t_{k_l}}}. \tag{3.55}$$

Conversely, if $M_{t_{k_l}} > M^*_{t_k}$, agents will place limit ask orders to short their position in risky asset investment, and the corresponding order price $a_{t_{k_l}}$ is denoted as

$$a_{t_{k_l}} \sim U(P_{t_k}, P_{t_k} e^{(\beta_{t_k} - \frac{1}{2}\sigma^2)\Delta t + \sigma\xi\sqrt{\Delta t}}), \tag{3.56}$$

where $\xi \sim N(0,1)$ is a standard normal variable and the time step $\Delta t = 1$. While the limit order volume $h_{t_{k_l}}$ is presented as

$$h_{t_{k_l}} = \frac{M_{t_k} - M^*_{t_{k_l}}}{a_{t_{k_l}}}. \tag{3.57}$$

3.7.2 CRRA Case

In initial, agent i generates his first belief β_0 about the value of μ with a Gaussian distribution $\beta_0 \sim \mathbf{N}\left(\mu_{\beta_0}, \sigma^2_{\beta_0}\right)$. Therefore, his optimal strategy using the CRRA utility function is to hold the θ^*_0 percentage of risky assets in the first period of the portfolio. Since no credit purchase and sell is allowed in the system, the agent's optimal investment in the risky asset is restricted to $\theta^*_t \in [0,1]$.

Let P_{t_k} represent the opening price at period t_k. During the period, if $\theta_{t_{k_l}} < \theta^*_{t_k}$, where $\theta^*_{t_k}$ is the position the agent wishes to be at the end of the period, the agent will place limit bid orders in order to increase their proportion of investment in a risky asset at the following intra-day periods. Then we assume that the limit bid order price $b_{t_{k_l}}$ is generated by

$$b_{t_{k_l}} \sim U(P_{t_k}, P_{t_k} e^{\beta_{t_k} - \frac{1}{2}\sigma^2 + \sigma\xi}), \tag{3.58}$$

where $\xi \sim N(0,1)$ is a standard normal variable.

In order to reach the optimal proportion position $\theta^*_{t_k}$, an agent is assumed to place a limit bid order to buy $h_{t_{k_l}}$ shares at price $b_{t_{k_l}}$. Once the order is execute completely, the value of risky asset in agent's portfolio is $\left(S_{t_{k_l}} + h_{t_{k_l}}\right) b_{t_{k_l}}$. At the same time, the value of portfolio is consisted of two parts: (1) the value of cash $V_{t_{k_l}}\left(1 - \theta_{t_{k_l}}\right)$, and (2) the value of risky asset $S_{t_{k_l}} b_{t_{k_l}}$. Then the limit bid order volume $h_{t_{k_l}}$ can be obtained

by solving the following equation

$$\frac{\left(S_{t_{k_l}} + h_{t_{k_l}}\right) b_{t_{k_l}}}{V_{t_{k_l}}\left(1 - \theta_{t_{k_l}}\right) + S_{t_{k_l}} b_{t_{k_l}}} = \theta^*_{t_k}, \tag{3.59}$$

and with a simple transformation, we can have

$$h_{t_{k_l}} = \left(\theta^*_{t_k} - 1\right) S_{t_{k_l}} + \frac{V_{t_{k_l}} \theta^*_{t_k}\left(1 - \theta_{t_{k_l}}\right)}{b_{t_{k_l}}}. \tag{3.60}$$

Similarly, if $\theta_{t_{k_l}} > \theta^*_{t_k}$, agents will place limit ask orders in order to decrease their proportion of investment in risky asset, and the corresponding order price $a_{t_{k_l}}$ and volume $h_{t_{k_l}}$ are written as

$$\begin{aligned} a_{t_{k_l}} &\sim U(P_{t_k}, P_{t_k} e^{\beta_{t_k} - \frac{1}{2}\sigma^2 + \sigma\xi}) \\ h_{t_{k_l}} &= \left(1 - \theta^*_{t_k}\right) S_{t_{k_l}} - \frac{V_{t_{k_l}} \theta^*_{t_k}\left(1 - \theta_{t_{k_l}}\right)}{a_{t_{k_l}}}, \end{aligned} \tag{3.61}$$

where where $\xi \sim N(0,1)$ is a standard normal variable. This formation of order price and volume indicates that the agents will only invest in a risky asset if they hold positive views of asset returns.

3.8 Illustration for Bayesian Learning

Now we introduce the framework for the Bayesian learning procedure. We assume that all agents believe a Geometric Brownian Motion dynamic of the stock price

$$d \log P_t = \mu P_t dt + \sigma P_t dW_t, \tag{3.62}$$

where $(W_t)_t$ is Wiener process.

Then the dynamic of β_t (3.5) can be transformed into

$$\begin{aligned} d\beta_t &= g(t)(\mu - \beta_t)dt + \sigma g(t) dW_t \\ &= g(t)(\mu dt + \sigma dW_t) - g(t)\beta_t dt \\ &= g(t) d\log P - g(t)\beta_t dt \end{aligned} \tag{3.63}$$

where g is defined by

$$g : t \in R_+ \mapsto \frac{v_0^2}{\sigma^2 + v_0^2 t}. \tag{3.64}$$

In order to embed the Bayesian learning in the model, the Equation (3.63) is discretized and approximated with the following steps

$$\begin{aligned}\beta_t &= \beta_{t-1} + \int_{t-1}^{t} d\beta_s \\ &= \beta_{t-1} + \int_{t-1}^{t} (g(s)d\log P - g(s)\beta_s ds) \\ &\approx \beta_{t-1} + g(t-1)\log\frac{P_t}{P_{t-1}} - g(t-1)\beta_{t-1}dt.\end{aligned} \tag{3.65}$$

In particular, agents can have learning beliefs of β_t by observing the historical stock price using Equation (3.65). To better illustrate the learning process of β_t, we present a simple example in Figure 3.1. The upper panel shows a stock path $\{P_t\}_t$ generated numerically by formula

$$P_t = P_{t-1} + \mu P_{t-1}dt + \sigma P_{t-1}\sqrt{dt}\xi, \tag{3.66}$$

where $\xi \sim N(0,1)$ is a standard normal variable. Given an initial price $P_0 = 100$, there are 1 000 price points calculated with a time step size $dt = 0.01$ and terminal time $T = 10$. The parameter μ and σ are set as 0 and 0.05 respectively.

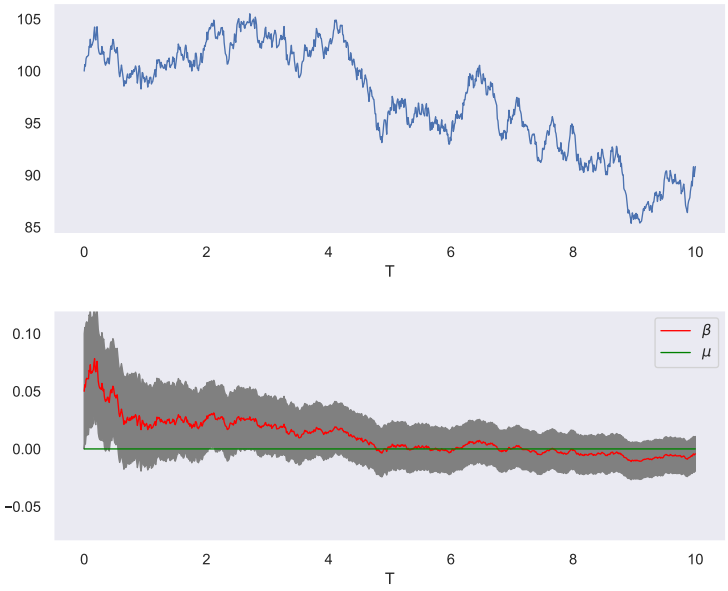


Figure 3.1: A simple example of stock path generated by Geometric Brownian Motion and the corresponding Bayesian updation of β_t.

The bottom panel shows the Bayesian learning process of β_t with an initial value $\beta_0 = 0.05$ and $v_0 = 0.05$. At the time t, the new β_t is calculated in Equation (3.65) by observing the price path in the upper panel. The grey area represents the range within one standard deviation of β_t, while the green straight line stands for the true value of μ. Initially, β_0 deviates from the actual value of μ. However, as more price information is obtained, the learning algorithm gets β_t closer to value of μ. Meanwhile, the standard deviation of β_t narrows over the learning process, which indicates we are more confident that the value of β_t reveals the actual value of μ.

3.9 Timeline of the Events

In previous sections, we have mathematically formalised the Bayesian learning technique and the optimal strategy of agents. However, the event sequence is also vital to make an identical model for repeatable simulation results.

The model starts with $N^{learning}$ number of learning agents, whose initial beliefs β_0 about the value of μ are drawn from a normal distribution $N \sim (\mu_\beta, \sigma_\beta^2)$. Each agent has a portfolio consisting of C_0 cash and S_0 shares at the beginning.

The following list describes the sequence of events that happen at the level of period and corresponding intra-day period

1. At the beginning of the each period, agents observe the historical stock price and update their belief β_t about the value of μ using Equation (3.65). Then their optimal strategies are determined in the following ways

    - If agents have a CARA utility function, they calculate their optimal risky asset investment M_t^* using Equation (3.26).
    - If agents have a CRRA utility function, they calculate their optimal proportion of risky asset θ_t^* using Equation (3.50) and (3.52).

2. For each intra-day period within a period, agents check their existing limit orders and withdraw them if the orders exist beyond one period. Then one agent is selected randomly to place a limit order introduced in Section 3.7.

3.10 Simulation

In the beginning, we present a pseudocode of an ABM simulation with trend-following agents in CRRA case to illustrate how we can obtain outputs based on the math formulas introduced in previous sections.

Algorithm 2 The pseudocode of an ABM simulation in CRRA case

Require: $N^{learning},\ T,\ L\ \in \mathbb{N}^+$
  $t \leftarrow 1$
  $l \leftarrow 0$
  $\gamma \leftarrow 1$
  $P_0 \leftarrow p_0$
  $\beta_0^i \leftarrow \mathbf{N}\left(\mu_{\beta_0}, \sigma_{\beta_0}^2\right),\ for\ i = 1, 2, \ldots, N^{learning}$
  set up the initial portfolio account for each agent
  while $t \leq T$ **do**
    while $l \leq L$ **do**
      if $l = 0$ **then**
        if Agents have learning believes **then**
          Update agents' believes using Equation (3.65)
        end if
        $l \leftarrow l + 1$
      else if $1 \leq l \leq L$ **then**
        randomly select an agent
        calculate his current and optimal risky asset proportion θ and θ^* (3.52)
        if $\theta < \theta^*$ **then**
          place a limit bid order using Equation (3.58) and (3.60)
        else if $\theta > \theta^*$ **then**
          place a limit ask order using Equation (3.61)
        end if
        update the new market price P_{t_l} following the rules in Section 2.4
        $l \leftarrow l + 1$
      end if
    end while
    $t \leftarrow t + 1$
  end while
  output a set of simulated price $\{P_{1_1}, P_{1_2}, \ldots, P_{T_L}\}$

In specfic, the population of learning agents $N^{learning}$ is set as 100. Each agent is given $S_0 = 100$ shares and $C_0 = 10\,000$ cash at the beginning. Each simulation consists of 2 000 periods, within which has 300 intra-day periods. The initial stock price is $P_0 = 100$ and the risk-free rate is assumed as $r_f = 0$.

In the following section, we present various results of simulations to demonstrate the model's inherent properties. In the first two subsections, we hold β_t fixed as a

constant, illustrating the scenario where these agents do not learn at all.

3.10.1 Fixed Belief about Value of Drift in CRRA Case: Agents Have the Same Proportion of Initial Investment in Risky Asset

We start from the simplest model, in which each agent has a fixed individual belief about the value of μ throughout the simulation. All agents are assumed to have a CRRA utility function with their risk aversion parameter $\gamma = 1$. This simplification aims to investigate how the agent belief distribution and their optimal strategy affect the dynamic of an asset price.

At the beginning of the simulation, agents start with a randomly held belief β_0 about the value of price drift μ sampled from a normal distribution $N \sim (0.005, 0.001^2)$, and their beliefs will remain the same over the simulation. Meanwhile, all agents are assumed to have the same knowledge of the price volatility $\sigma = 0.1$. Here we define an expected optimal risky asset proportion for each agent

$$\tilde{\theta}^* = E\left[\theta_t^*\right] = \frac{E\left[\beta_t\right]}{\sigma^2}, \tag{3.67}$$

which is an expectation of agent optimal strategy in Equation (3.52). Since all agents hold fixed beliefs, they will equalise their investment in both risky and risk-free asset over the simulation given $\tilde{\theta}^* = 0.5$.

It is noteworthy that the normal distribution for generating the agent beliefs should have a positive mean since agents are only willing to invest in the risky asset if they believe there will be a positive return or they will choose to hold risk-free money. Meanwhile, the diversity of the belief about the asset return among agents is essential for the model initialisation. If all agents have the same belief, there will be no difference between the bid and ask prices, and no transaction will be executed. Therefore the standard deviation of the agent belief distribution should be large enough to create price discrimination.

Figure 3.2 presents two simulation results consisting of two stock paths and their corresponding histograms of agents optimal strategy θ^*. Given that the initial proportion of risky investment for each agent θ_0 is set as 0.5, the path in the upper panel of

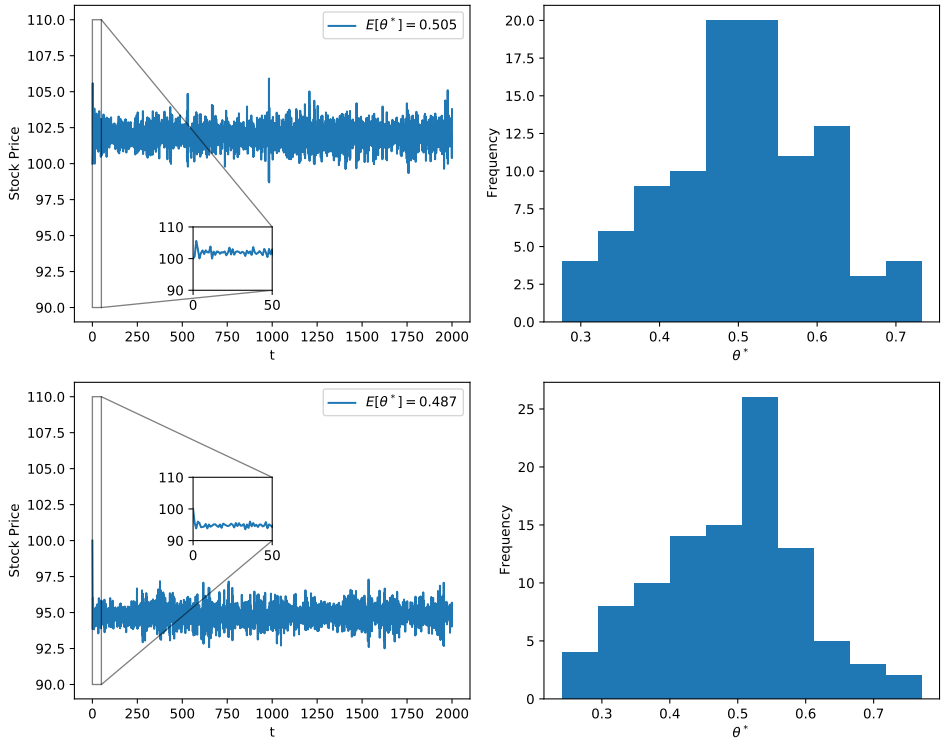


Figure 3.2: Upper panel: A price sample path and the corresponding histogram of agents optimal strategy θ^* with $\tilde{\theta}^* = 0.505$. Bottom Panel: A price path sample and the corresponding histogram of agents optimal strategy θ^* with $\tilde{\theta}^* = 0.487$.

Figure 3.2 shows a surge in stock price at the first periods, where the expected optimal strategy among agents $\tilde{\theta}^*$ equals to 0.505. It indicates that agents are more likely to place the bid orders during the first few periods. Conversely, the path in the bottom panel of Figure 3.2 displays a rapid drop of price at the first few periods, where $\tilde{\theta}^*$ equals to 0.487. It implies that the excess ask orders are placed by agents since the expected optimal strategy among agents is lower than their initial investment in a risky asset.

Four additional sample paths are displayed in Figure 3.3, in which we can observe that when $\tilde{\theta}^* > \theta_0$, the orange and blue sample paths move upward and the magnitude is positively correlated to the distance between $\tilde{\theta}^*$ and θ_0. Similarly, when $\tilde{\theta}^* < \theta_0$, the red and green sample paths drops quickly at the first period. All four sample paths present a stable movement in the rest periods.

Figure 3.4 shows the corresponding bid-ask spread of four sample paths. During the first few periods, the bid-ask spreads fluctuated in the range $(0, 2.5]$ and their price series show deterministic upward or downward trends during these periods. After that,

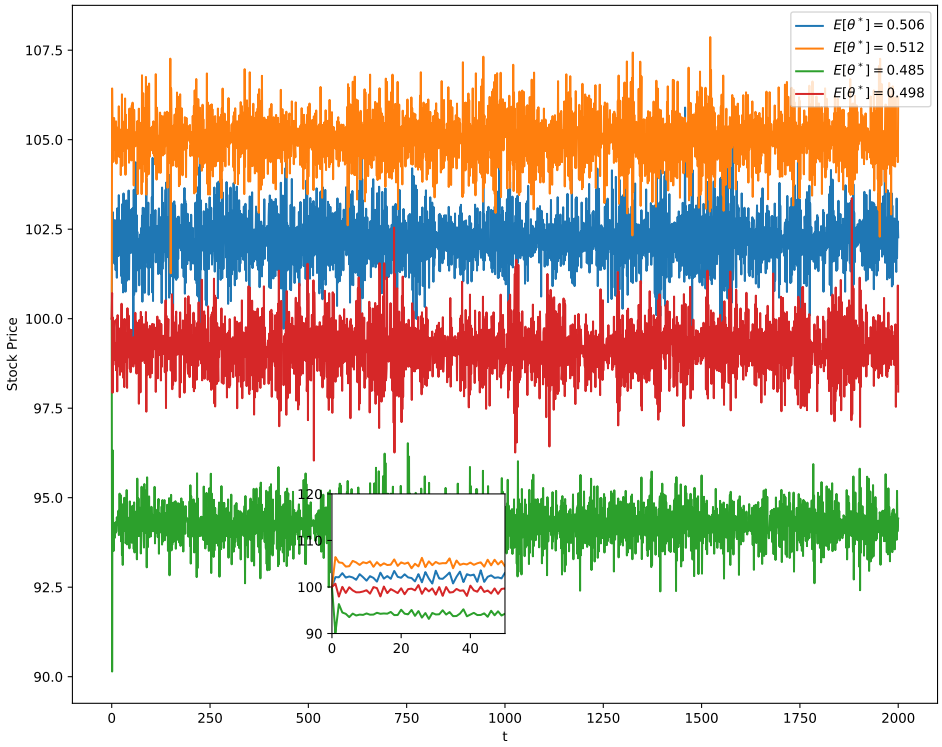


Figure 3.3: Four sample paths simulated under the condition that agents have fix beliefs about the value asset drift μ, where the belief is sampled from $N \sim (0.005, 0.001^2)$.

the bid-ask spreads move stably in the range $(0, 4]$ and the price series show a mean reversion pattern simultaneously.

Given that agents have fixed beliefs sampled from a normal distribution $N \sim (0.005, 0.001^2)$, we can observe the deterministic and stationary pattern of the sample path. At the same time, the bid-ask spread under the deterministic pattern is smaller and narrower than the stationary one. Moreover, the expected optimal strategy θ^* among agents determines the direction of price movement at the first few periods. When $\tilde{\theta}^* > \theta_0$, it causes an upward trend of sample-path. Conversely, a downward trend of sample-path is determined by the condition that $\tilde{\theta}^* < \theta_0$.

When we choose different standard deviations for generating agents fixed beliefs and keep the other parameters unchanged, a long term deterministic trend occurs in the simulation results.

In Figure 3.5, the upper panel displays four sample paths simulated under the condition that agents have fixed beliefs about the value of asset drift μ. Each sample path has a unique standard deviation σ_{β_0} for generating agents beliefs with a normal distribution $N \sim \left(0.005, \sigma^2_{\beta_0}\right)$. While the bottom panel shows the first 200 periods of

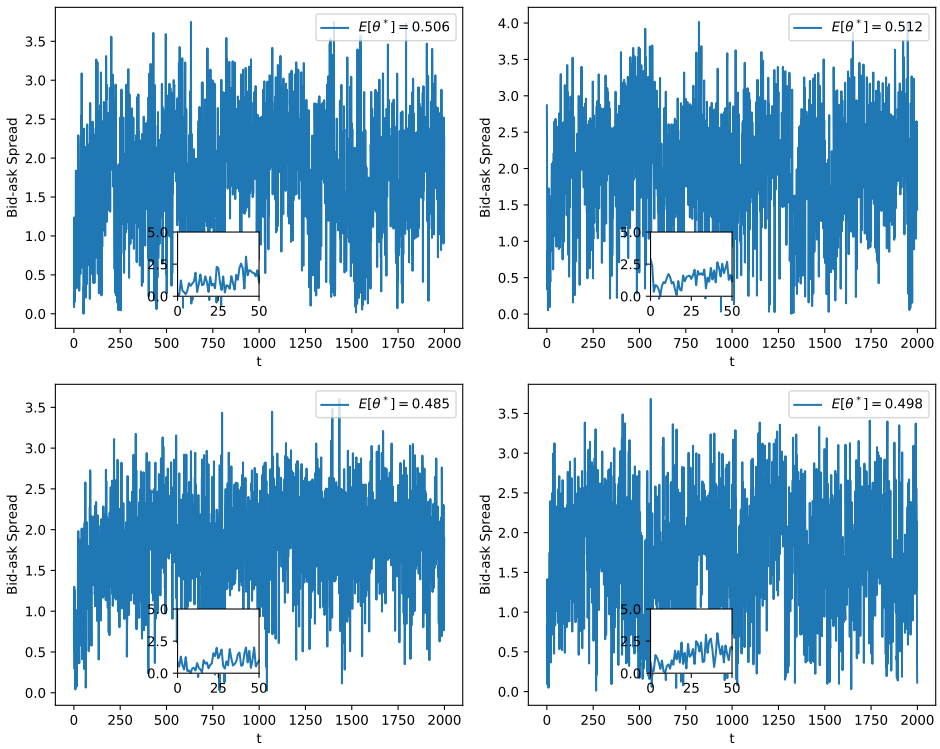


Figure 3.4: The corresponding bid-ask spreads of four sample paths in Figure (3.3).

the four sample paths in upper panel. As σ_{β_0} increasing, it causes a larger variation in stock price. The red sample path which has $\tilde{\theta}^* = 0.543$ and $\sigma_{\beta_0} = 0.004$ increases by 160 at the first 10 periods due to $\tilde{\theta}^* > \theta_0 = 0.5$. However, the stationary pattern of sample path in the case that $\sigma_{\beta_0} = 0.001$ is replaced by a deterministic downward trend in the red sample path after the first 10 periods. For the other three sample paths, it is hard to verify whether there is a stationary or deterministic pattern after the first few periods.

Figure 3.6 presents the corresponding bid-ask spreads of four sample paths in Figure 3.5. When σ_{β_0} increases, the bid-ask spread of sample path shows a larger variation. If we set the bid-ask spread with $\sigma_{\beta_0} = 0.001$ as the baseline, the case with $\sigma_{\beta_0} = 0.002$ shows no significant difference comparing with the baseline, but has a slight higher maximum value. For the case has $\sigma_{\beta_0} = 0.003$, although the range of bid-ask spread triples, it will periodically decrease to the similar range of the baseline. The case with $\sigma_{\beta_0} = 0.004$ has a similar pattern with the case $\sigma_{\beta_0} = 0.003$, which has a more extreme maximum value up to 20 and the a decreasing range.

Figure 3.7 displays the first 200 periods of bid-ask spread in Figure 3.6, in which we can observe that for cases $\sigma_{\beta_0} = 0.002$ and $\sigma_{\beta_0} = 0.001$, the bid-ask spreads fall in

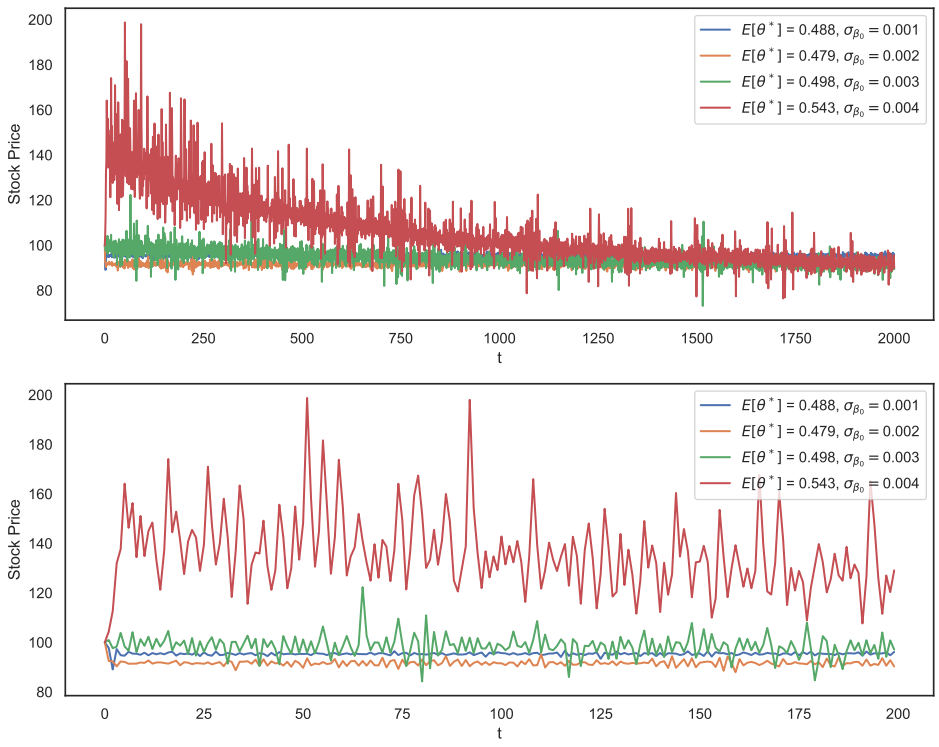


Figure 3.5: Upper panel: Four sample paths simulated under the condition that agents have fixed beliefs about the value of asset drift μ. Each sample-path has a unique standard deviation σ_{β_0} for generating agents beliefs with a normal distribution $N \sim \left(0.005, \sigma^2_{\beta_0}\right)$. Bottom panel: The first 200 periods of four sample paths showed in the upper panel.

range $[0, 2.5]$ accompanying with a deterministic trend in price at the first 10 periods. While in cases $\sigma_{\beta_0} = 0.003$ and $\sigma_{\beta_0} = 0.003$, larges fluctuations in price are more likely to come with the corresponding bid-ask spreads surpassing the range $[0, 4]$.

By shedding lights on the bid-ask spread of sample paths with different values of σ_{β_0}, we can roughly deduce that the stationary pattern of the stock price results from the bid-ask spread varying stably in the range $[0, 4]$. The variation of the bid-ask spread in a narrower or broader range than $[0, 4]$ leads to a deterministic trend in price, which usually happens at the first ten periods of simulation and the case with $\sigma_{\beta_0} = 0.004$.

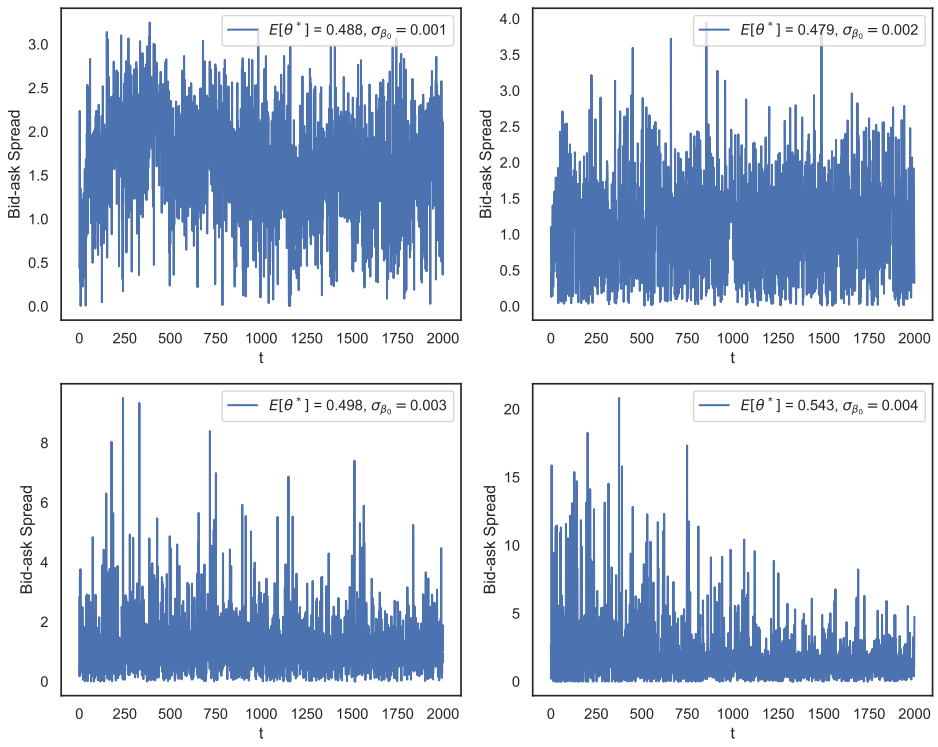


Figure 3.6: The corresponding bid-ask spreads of four sample paths in Figure 3.5.

3.10.2 Fixed Belief about the Value of Drift in CRRA: Agents Have the Different Initial Proportion of Investment in Risky Asset

In the previous section, we assume that each agent's initial proportion of risky asset investment equals 0.5. A deterministic trend of the stock price at the first period is caused by the difference between expected optimal strategy among agents $\tilde{\theta}^*$ and their initial proportions of risky asset investment θ_0.

Here we change the initial setting of agents' portfolios,

- A number of 30 learning agents are allocated with $S_0 = 50$ shares and $C_0 = 15\,000$ cash.
- A number of 30 learning agents are allocated with $S_0 = 150$ shares and $S_0 = 5\,000$ cash.
- A number of 40 learning agents are allocated $S_0 = 100$ shares and $10\,000$ cash.

The population of learning agents remains the same $N^{learning} = 100$. Given the initial stock price $P_0 = 100$, the total amount of cash and share in the system stays

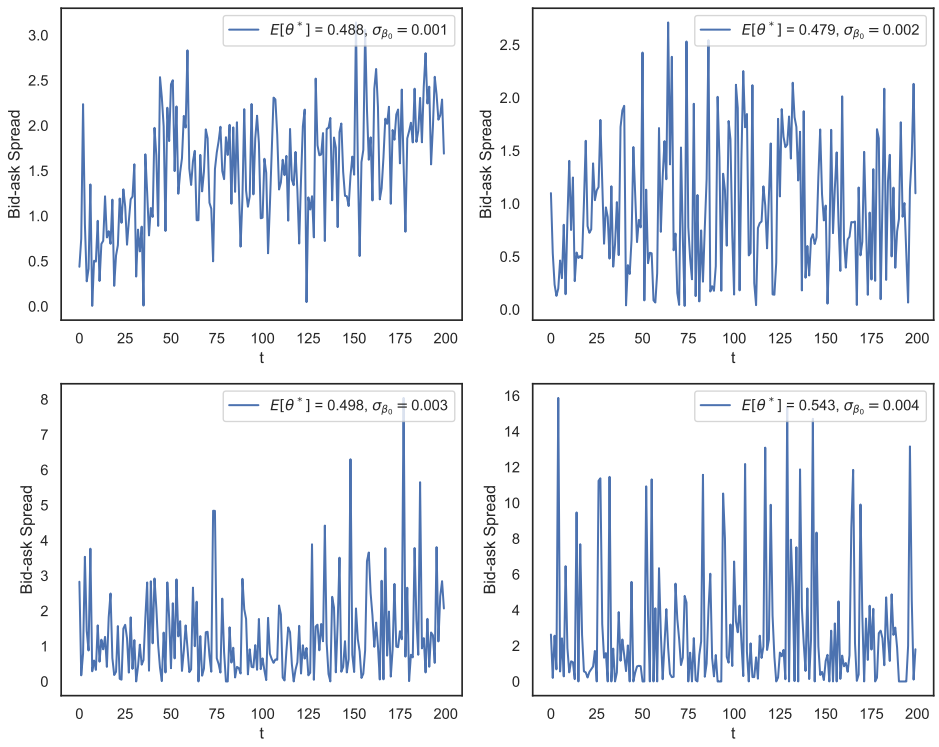


Figure 3.7: The first 200 periods of bid-ask spreads corresponds to four sample paths in Figure 3.5.

the same with the last section, but the initial proportion of risky asset investment of agents is divided into three levels: 25%, 50% and 75%. Under the new setting, each agent has the same initial wealth at the beginning of the simulation, and the other parameters will remain unchanged.

Figure 3.8 illustrates four sample paths simulated under the condition that agents have fixed beliefs about the value of asset drift μ and three proportion levels of initial risky asset investment, where the belief is sampled from $N \sim (0.005, 0.001^2)$. This result has no significant difference from the simulation because each agent has the same initial risky asset investment proportion. The sample paths display a deterministic trend at the first few periods, followed by a mean reversion pattern.

In Figure 3.9, the corresponding bid-ask spreads mostly fall in the range $[0, 2.5]$ at the first 10 periods, after that the bid-ask spreads vary in the range $[0, 5]$. Although agents have different proportions of initial risky asset investments, it does not change the pattern of sample-path under the condition that agents beliefs are a sample from $N \sim (0.001, 0.001^2)$. It indicates that the distribution of agents beliefs and agents optimal strategy are two essential factors determining the price pattern.

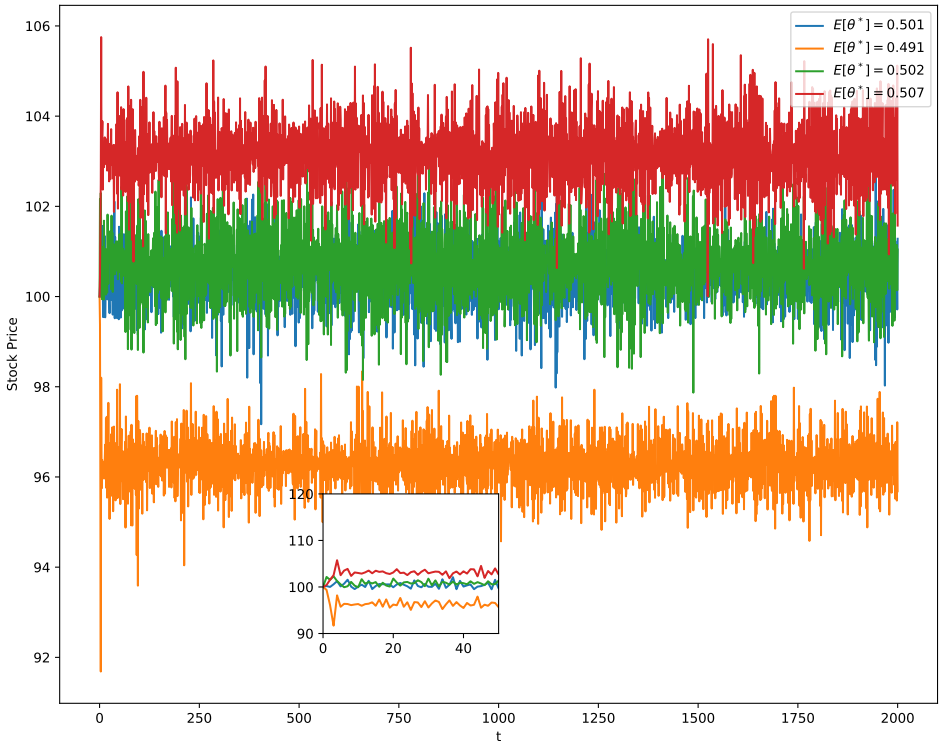


Figure 3.8: Four sample paths are simulated under the condition that agents have fixed beliefs about the value of asset drift μ and three proportion levels of initial risky asset investment, where the belief is sampled from $N \sim (0.005, 0.001^2)$.

Meanwhile, in Figure 3.10, the upper panel shows four sample paths simulated under the condition that agents have fixed beliefs about the value of asset drift μ and three proportion levels of initial risky asset investment. Each sample-path has a unique standard deviation σ_{β_0} for generating agents beliefs with a normal distribution $N \sim \left(0.005, \sigma_{\beta_0}^2\right)$. While the bottom panel in Figure 3.10 displays the first 200 periods of four sample paths in the upper panel.

The cases that $\sigma_{\beta_0} = 0.001$ and $\sigma_{\beta_0} = 0.002$ share the same pattern with a deterministic trend at the first 10 periods and a mean reversion pattern coming after. However, the cases that $\sigma_{\beta_0} = 0.003$ and $\sigma_{\beta_0} = 0.004$ both display a deterministic trend throughout the whole simulation.

The corresponding bid-ask spreads displayed in Figure 3.11 and Figure 3.12 provide us with more details that an unstable range of bid-ask spread is more likely to contribute to a deterministic trend in asset price. More precisely, the case $\sigma_{\beta_0} = 0.001$, the bid-ask spreads fall in a stable range $[0, 4]$, which shows a mean reversion pattern in price path after the first few periods. While, the cases $\sigma_{\beta_0} = 0.003$ and $\sigma_{\beta_0} = 0.004$

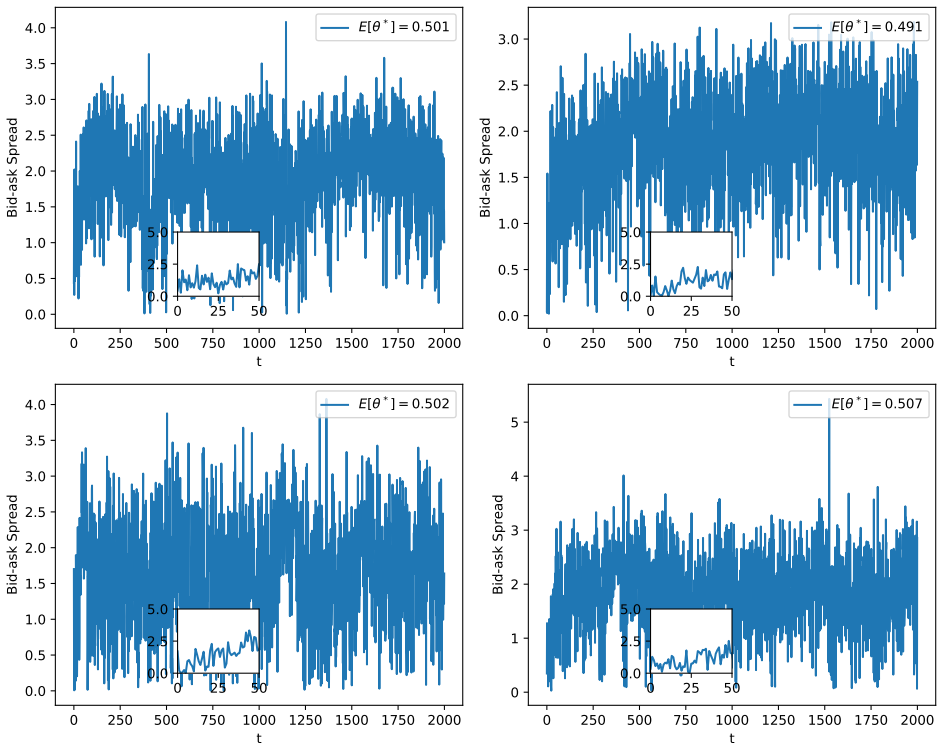


Figure 3.9: The corresponding bid-ask spreads of four sample paths in Figure 3.8.

have unstable ranges of bid-ask spread, which result in a deterministic pattern in asset price.

3.10.3 Leaning Belief about the Value of Drift in CRRA Case: Agents Have the Same Proportion of Initial Investment in Risky Asset

In this part, we assume that all agents have a CRRA utility function with their risk aversion parameter $\gamma = 1$. At the beginning of the simulation, agents generate their first beliefs β_0 about the value of price drift μ, sampled from a normal distribution $N \sim (0.005, 0.001^2)$. They will update their beliefs using Bayesian learning (See Section 3.8) at the beginning of each period. Meanwhile, each agent has the same standard deviation v_0 of a Gaussian prior distribution. We will investigate the pattern of stock price by choosing different values of v_0. All other parameters will remain unchanged.

In Figure 3.13, these four sample paths are presented under the condition that agents have learning beliefs about the value of μ and $v_0 = 0.003$. Four sample paths

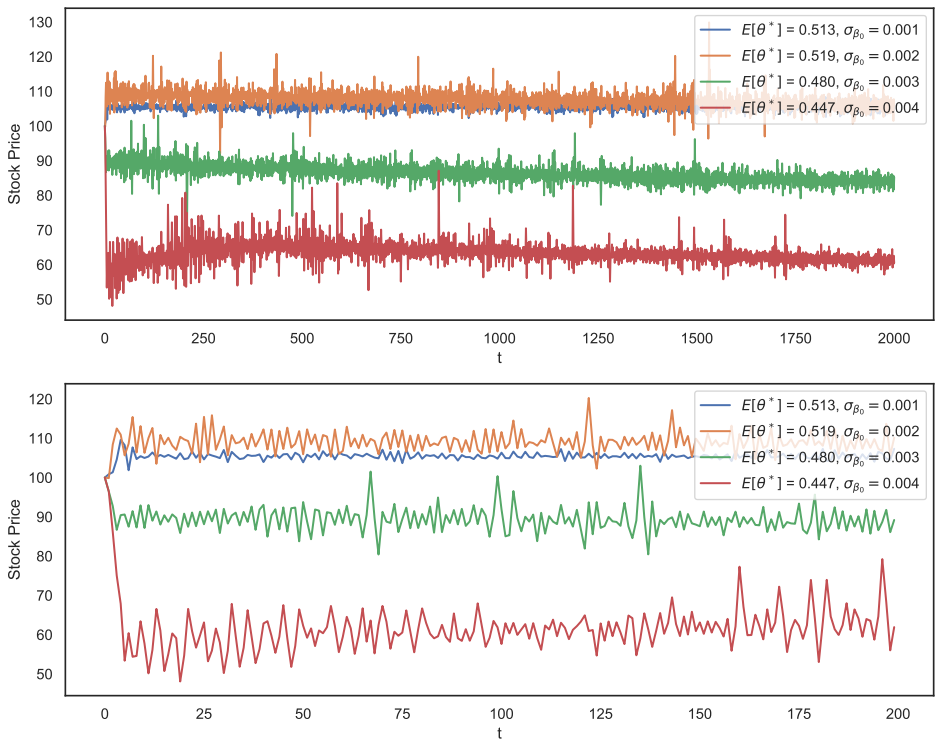


Figure 3.10: Upper panel: Four sample paths simulated under the condition that agents have fixed beliefs about the value of asset drift μ and three proportion levels of initial risky asset investment. Each sample-path has a unique standard deviation σ_{β_0} for generating agents beliefs with a normal distribution $N \sim \left(0.005, \sigma^2_{\beta_0}\right)$. Bottom panel: The first 200 periods of four sample paths showed in the upper panel.

all have a downward trend with a decreasing gradient. The expected optimal strategy among agents for each sample path at the terminal period is displayed at the right-up corner.

Given that agents' first beliefs about the value μ are sampled from $N \sim (0.005, 0.001^2)$, agents expect the stock to have a return around 0.005 for the next period. For the orange and green sample paths, agents are more likely to place the bid orders instead of the ask orders at the beginning of the simulation, which causes the price goes to up. Consequently, the proportion of risky asset investment increases for agents who place bid orders, implying that these agents expect a higher stock return. However, under the setting that all agents have limited and shares and no credit transaction is allowed in the model, agents will run out of liquidity even they want to place bid orders in the market. Then the increment of a stock price is more minor agents' expectations, and agents will downgrade their beliefs about the value of μ. The majority of agents who hold opposing opinions on stock return will accelerate the fall of stock price, making

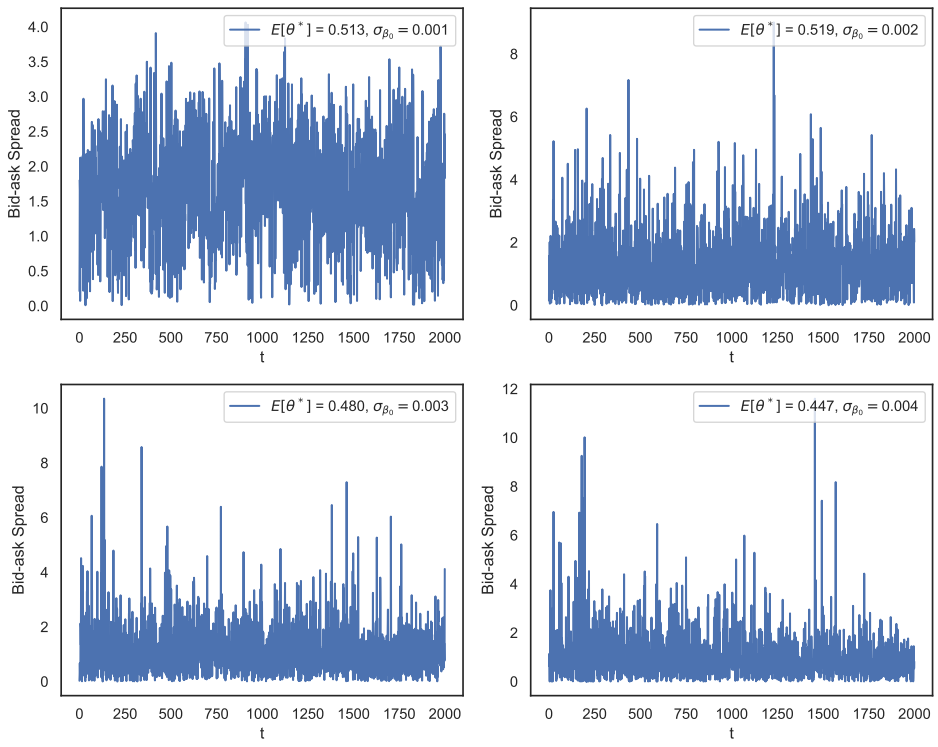


Figure 3.11: The corresponding bid-ask spreads of four sample paths in Figure 3.10.

agents lower their expectations. A negative feedback loop is created in this way for the blue sample-path; due to more agents operating to decrease their investment in a risky asset, the market enters into a negative feedback loop. Therefore, the expected optimal strategy among agents for four sample paths is far from their initial positions with values around 0.11.

Figure 3.14 shows the bid-ask spreads of four sample paths in Figure 3.13, which decay quickly over time. Although agents have different initial beliefs, the Bayesian learning skill will reduce their belief differences. It is the reason that we observe a decreasing bid-ask spread. Meanwhile, it satisfies the pattern we state in Section 3.10.1 that an unstable range of bid-ask spread usually accompanies a deterministic trend in stock price.

In Figure 3.15, we change the standard deviation v_0 of agents Gaussian prior distribution to 0.004, which indicates agents are less confident about their first beliefs β_0 and will amplify the adjustment of beliefs through Bayesian updating. Four sample paths show a dramatic price drop and followed by a horizontal movement. The expected optimal strategy among agents are all negative and near 0. The negative θ^* means that agents are willing to borrow the risky asset and then sell it for cash to

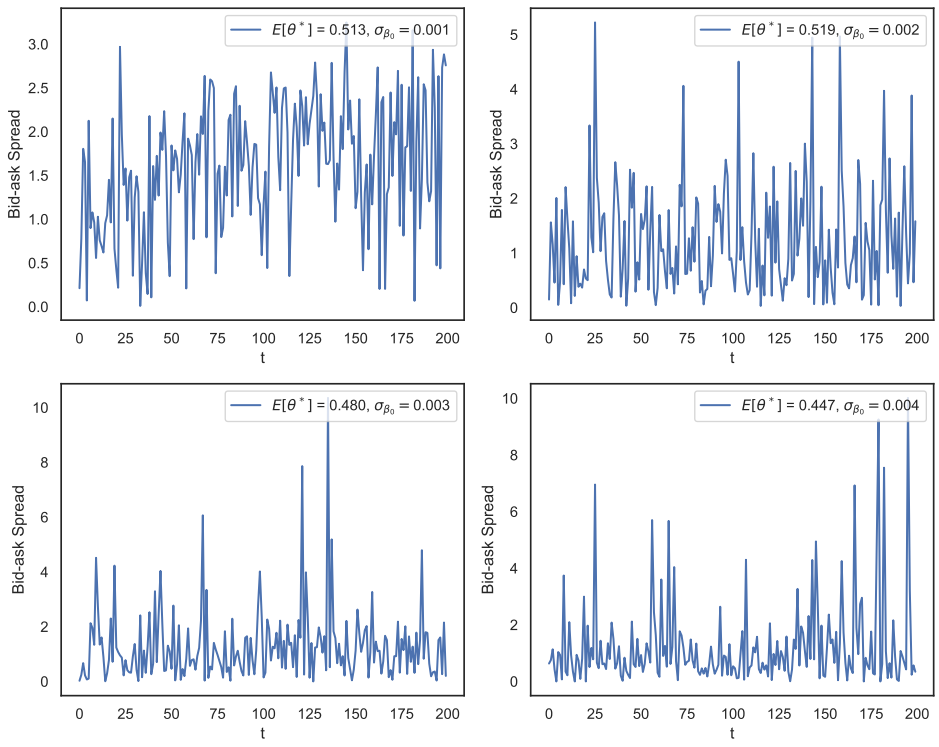


Figure 3.12: The first 200 periods of bid-ask spreads corresponds to four sample paths in Figure 3.10.

earn the risk-free interest. However, credit transaction is not allowed in the model, and only the ask orders appear in the limit order book, which explains that the stock price is flat after 250 periods.

Figure 3.16 illustrates that bid-ask spreads under $v_0 = 0.004$ condition decay more faster than $v_0 = 0.003$ and it turns to 0 after no bid orders occur in the market.

3.11 Parameter Sensitivity Analysis

In the previous section, we have illustrated some model outputs under several scenarios. However, the parameter sensitivity analysis is still vital to explain how the parameter changes influence model outcomes, such as price, log-return, bid-ask spread, and optimal strategy. This section will implement a parameter sensitivity analysis based on the one-factor-at-a-time (OFAT) method, which changes one parameter at a time and remains the other parameters constant. It can allow us to decide how the initial condition of parameters should be chosen and how the summary statistics should be selected and linked to characterise the model's behaviours. Besides, it can identify

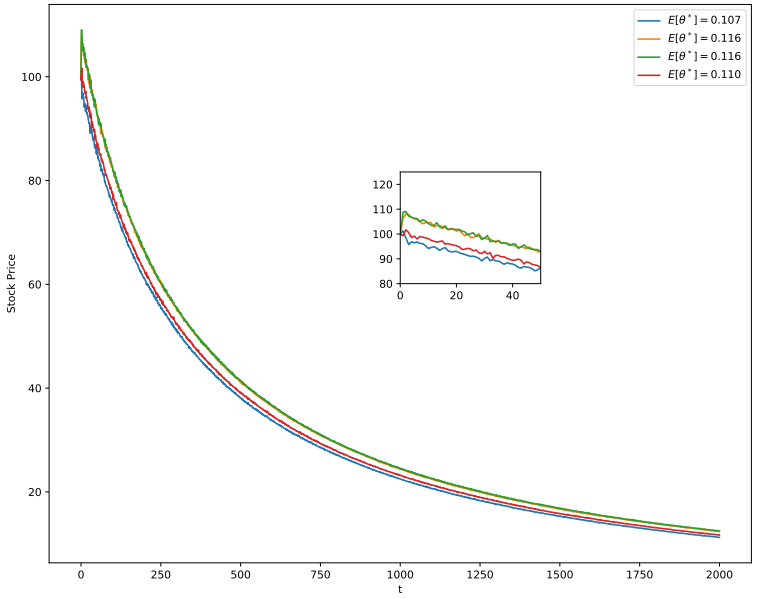


Figure 3.13: Four sample paths are simulated under the condition that agents have learning beliefs about the value of μ and $v_0 = 0.003$. The expected optimal strategy among agents for each sample path at the terminal period is displayed at the right-up corner.

two basic facts: (1) whether a linear relationship exists between the parameter and the output; (2) whether the tipping points exist that the output reacts vastly to a tiny parameter change.

The parameters we chosen to investigate are μ_β, σ_β and v_0 introduced in Section (3.9) and Equation (3.6). For μ_β and σ_β, they determine the distribution of agents' beliefs β_0 when the model simulation is initialised. A wrong choice of $(\mu_\beta, \sigma_\beta)$ may lead to two main problems: (1) Agents beliefs stack in a small range causing the maximum distance of limit order price smaller to than the minimum tick size of stock price; (2) Agents beliefs are widely spread that the best bid and best ask order will never crossover with each other to make a transaction. The parameter v_0 stands for the standard deviation of μ's prior distribution when $t = 0$, which affects the learning rate of β_t.

We are interested in a set of statistics $S = \{m_p, m_r, m_d, m_{\theta^*}, m_\theta\}$, which is the mean of the following model outputs: (1) simulated stock price $\{p_t, t = 1, \ldots, T\}$, (2) simulated log-return $\{r_t, t = 1, \ldots, T\}$, (3) simulated bid-ask spread $\{d_t, t = 1, \ldots, T\}$, (4) terminal optimal proportion of agents $\{\theta_i^*, i = 1, \ldots, N^{learning}\}$, and (5) terminal real proportion of agents $\{\theta_i, i = 1, \ldots, N^{learning}\}$. For each run $(h = 1, 2, \ldots, H)$,

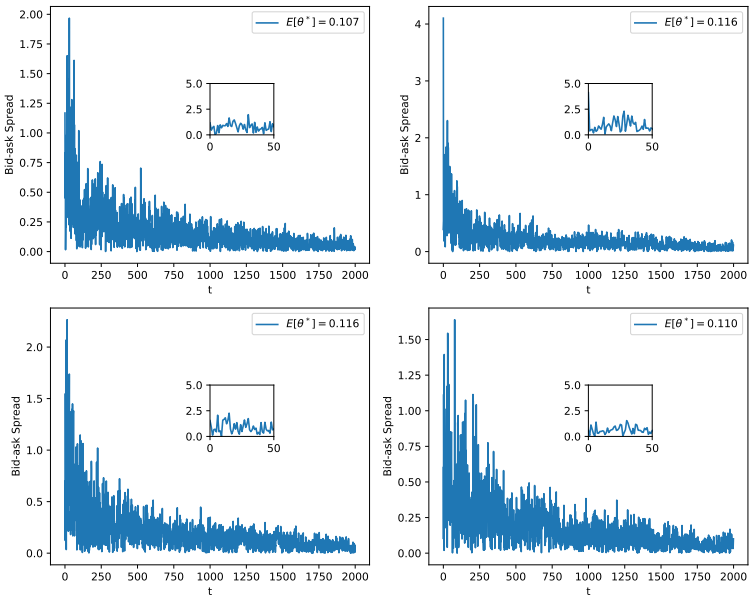


Figure 3.14: The corresponding bid-ask spread of sample paths shown in Figure 3.13.

the simulation will produce a sample of set $S^h = \{m_p^h, m_r^h, m_d^h, m_{\theta^*}^h, m_\theta^h\}$. Due to the stochastic nature of simulation, the element in S^h with be different for each independent run. Therefore, we implement H independent simulations to get $\{S^1, S^2, \ldots, S^H\}$, and denote the average performance of desired statistics S as

$$\begin{aligned}\tilde{S} &= \left\{ \frac{\sum_1^H m_p^h}{H}, \frac{\sum_1^H m_r^h}{H}, \frac{\sum_1^H m_d^h}{H}, \frac{\sum_1^H m_{\theta^*}^h}{H}, \frac{\sum_1^H m_\theta^h}{H} \right\} \\ &= \{\tilde{m}_p, \tilde{m}_r, \tilde{m}_d, \tilde{m}_{\theta^*}, \tilde{m}_\theta\}\,.\end{aligned} \tag{3.68}$$

Then, we will investigate how the $\tilde{S}$ changes against μ_β, σ_β and v_0 respectively.

In the following parts, we run $H = 50$ independent simulations for the sensitivity analysis. For each run of the simulation, there are $T = 200$ periods and each period consists of $L = 300$ intra-day periods. A population of $N^{learning} = 100$ learning agent are involved in the model. The initial stock price is set as $P_0 = 100$. Each agent is assume to have the same portfolio, which includes $C_0 = 10\,000$ cash and $S_0 = 100$ share. All agents have a CRRA utility function with $\gamma = 1$ and implement optimal portfolio strategy introduced in Section 3.5.

3.11.1 Model Outputs against Parameter μ_β

We firstly look at how the parameter μ_β influences the statistic $\tilde{S}$ in (3.68). The value of μ_β is selected evenly starting from 0.0035 to 0.0065 with a step size $\Delta = 0.0005$.

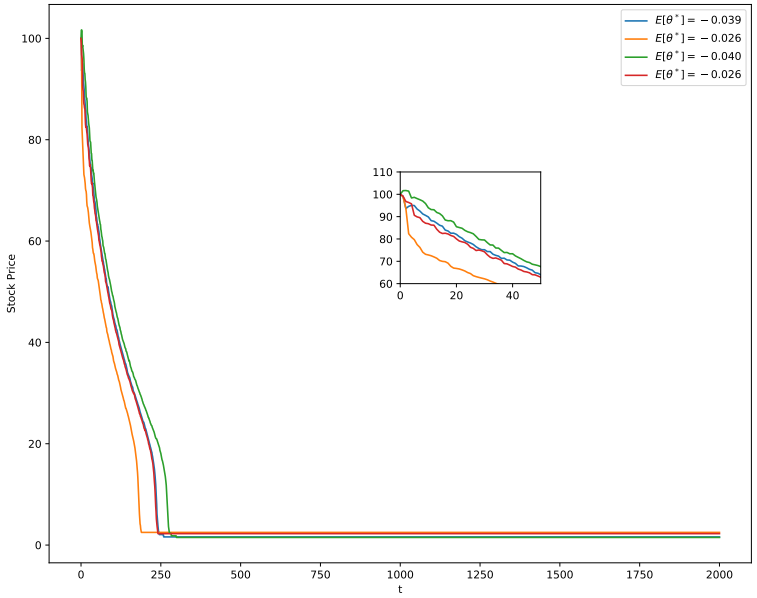


Figure 3.15: Four sample paths are simulated under the condition that agents have learning beliefs about the value of μ and $v_0 = 0.004$. The expected optimal strategy among agents for each sample path at the terminal period is displayed at the right-up corner.

While the value of σ_β is set as 0.005 and the parameter v_0 equals to 0.001. The other parameters remain the same as stated in Section (3.11).

Figure 3.17 roughly shows the distribution of each element in the set of statistics $\{m_p, m_r, m_d, m_{\theta^*}\}$ against parameter μ_β via the box plot. Each box and whisker plot contains $H = 50$ observations generated independently with the same parameter input, where the whisker is extended no more than $1.5 \times$ IQR (the interquartile range = 75th percentile - 25th percentile) from the edges of the box, ending at the farthest data point within that interval and outliers are plotted as separate dots.

The top-left panel displays the mean of the stock price $\{m_p^1, m_p^2, \ldots, m_p^H\}$ increases globally as the value of μ_β increases. However the spread of the stock price is wider at two edge points ($\mu_\beta = 0.0035$ and $\mu_\beta = 0.0065$) with more outliers. Similarly, for the mean of the bid-ask spread $\{m_d^1, m_d^2, \ldots, m_d^H\}$ and the optimal proportion $\{m_{\theta^*}^1, m_{\theta^*}^2, \ldots, m_{\theta^*}^H\}$, this positive relationship against the parameter μ_β can also be observed. While the mean of optimal proportion has a consistent pattern at two edge points. By contrast, the distribution of the mean of log-return $\{m_r^1, m_r^2, \ldots, m_p^H\}$ shows a stable pattern against the value of μ_β, but a wider distribution and more outliers exist at the two edge points.

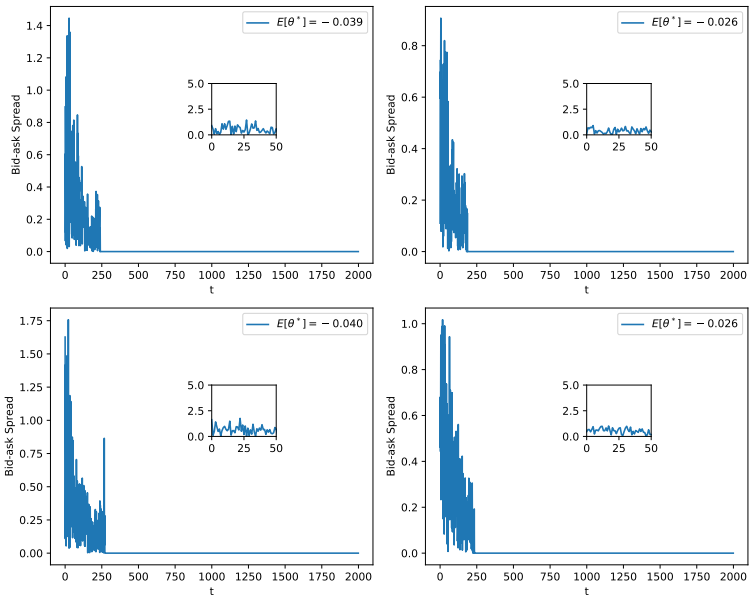


Figure 3.16: The corresponding bid-ask spread of sample paths shown in Figure 3.15.

The box plot presented in Figure (3.17) implies that given all other parameters fixed, the mean of model outputs $\{m_p, m_r, m_d, m_{\theta^*}\}$ can have a relatively traceable behaviour by choosing a suitable μ_β within the range $[0.0035, 0.0065]$. For the condition $\mu_\beta < 0.0035$, agents are more likely only to place ask orders since their initial risky asset proportions $\theta_0 = 0.5$ are greater than their optimal proportion θ^*, which will cause a large fluctuation in stock price and the corresponding log-return or even no counter-parties exist in the market. Likewise, for the condition $\mu_\beta > 0.0065$, agents are eager to submit bid orders to increase their risky asset proportion because of $\theta^* > \theta_0$. Meanwhile, a wider distribution of bid-ask spread occurs due to a high stock price, and for some cases, their bid-ask spread equal to 0 means no opposite order exists in the market.

Figure 3.18 illustrates how the change of parameter μ_β affects the average performance of desired statistic $S = \{m_p, m_r, m_d, m_{\theta^*}, m_\theta\}$. The mean of average stock price $\tilde{m}_p$ increases over the parameter μ_β and it shows a linear relationship when $\mu_\beta \geq 0.004$. For the mean of average log-return $\tilde{m}_r$, it is no significantly affected by the parameter μ_β, whose value bounces around 0. The bottom-left panel shows a positive linear relationship between the mean of average bid-ask spread $\tilde{m}_d$ with the parameter μ_β when $\mu_\beta < 0.055$. Then the slope of the curve becomes fluctuated. The mean of average optimal risky proportion of agents at the terminal time $\tilde{m}_{\theta_*}$ exactly has a linear

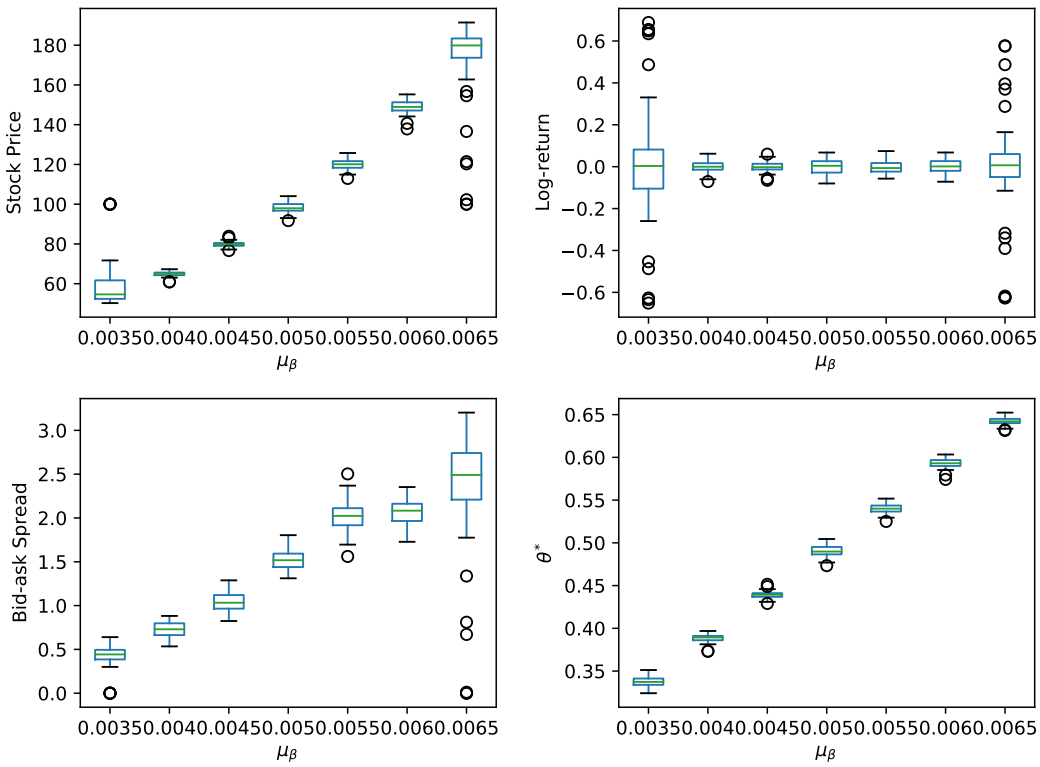


Figure 3.17: The box plots of statistics $\{m_p, m_r, m_d, m_{\theta^*}\}$ against the parameter μ_β.

relationship with the parameter μ_β and it overlaps with the mean of average real risky proportion of agents at the terminal time $\tilde{m}_\theta$ when $0.004 \leq \mu_\beta \leq 0.006$.

The sensitivity analysis of the model outputs against parameter μ_β suggests that given all other parameter setting as default, and we can choose the initial value of parameter μ_β from the range $[0.004, 0.006]$ to create a well-performed market, where all agents can achieve their optimal proportion strategy with efficient transactions at the end of the simulation. It avoids two problems in the simulation: (1)agents beliefs will locate in a small range causing the maximum distance of limit order price to be smaller than the minimum tick size of stock price when μ_β is very tiny; (2) agents beliefs will be widely distributed that the best bid and best ask order will never crossover with each other to make a transaction, when μ_β is relatively large. Meanwhile, we can extract features that the mean of average stock price and bid-ask spread has a positive linear relationship with μ_β in the range $[0.004, 0.006]$, which can be helpful when we calibrate our model against the real data.

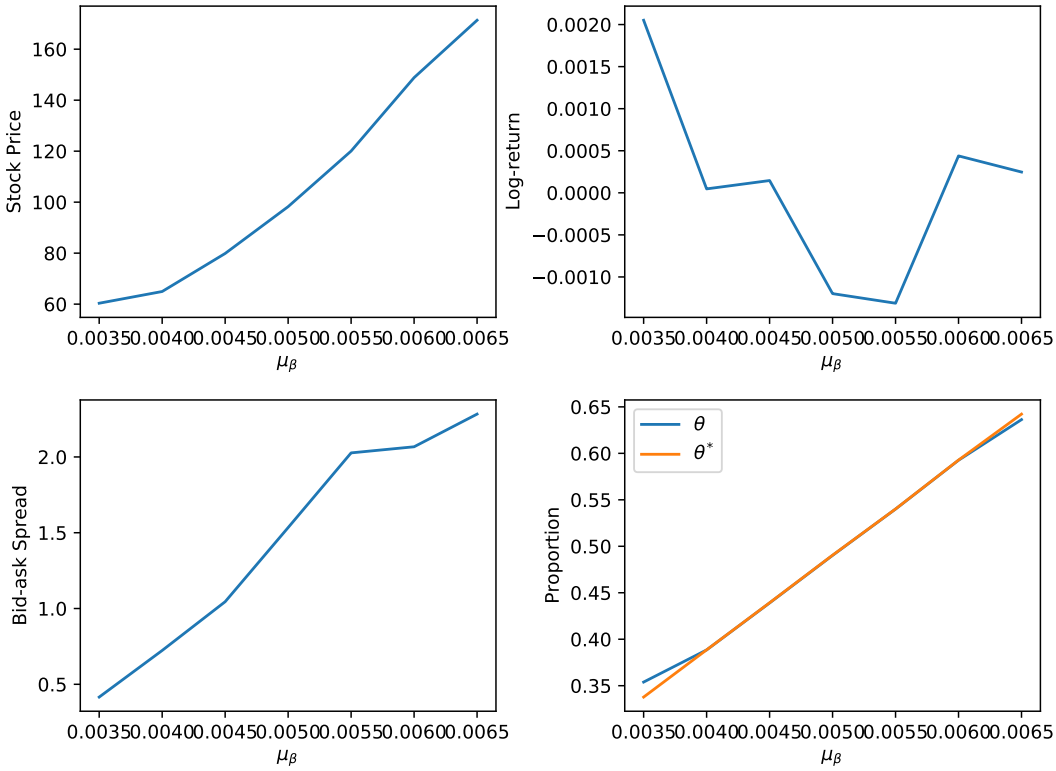


Figure 3.18: The sensitivity test of statistics $\{\tilde{m}_p, \tilde{m}_r, \tilde{m}_d, \tilde{m}_{\theta^*}, \tilde{m}_\theta\}$ against the parameter μ_β.

3.11.2 Model Outputs against Parameter σ_β

Here we look at how the parameter σ_β influences the statistic $\tilde{S}$ in (3.68). The value of σ_β is chosen equally starting from 0.0013 to 0.0027 with a step size $\Delta = 0.0004$. While the value of μ_β is set as 0.005 and the parameter v_0 equals to 0.001. The other parameters remain the same as stated in Section (3.11).

Figure 3.19 presents the box-plot of each element in a set of desired statistics $\{m_p, m_r, m_d, m_{\theta^*}\}$ against parameter σ_β. It can be directly viewed that the median value of $\{m_p^1, m_p^2, \ldots, m_p^H\}$, $\{m_r^1, m_r^2, \ldots, m_r^H\}$ and $\{m_{\theta^*}^1, m_{\theta^*}^2, \ldots, m_{\theta^*}^H\}$ has no significant change against the parameter σ_β, while the variation of them gradually increase as the value of σ_β increases. By contrast, the median value of $\{m_d^1, m_d^2, \ldots, m_d^H\}$ slightly decreases against σ_β and reaches the minimum at $\sigma_\beta = 0.002$. Then the median value of $\{m_d^1, m_d^2, \ldots, m_d^H\}$ slowly moves up with a wider spread.

Figure 3.20 displays how the change of parameter σ_β influences the average performance of desired statistics $S = \{m_p, m_r, m_d, m_{\theta^*}, m_\theta\}$. There is no linear relationship

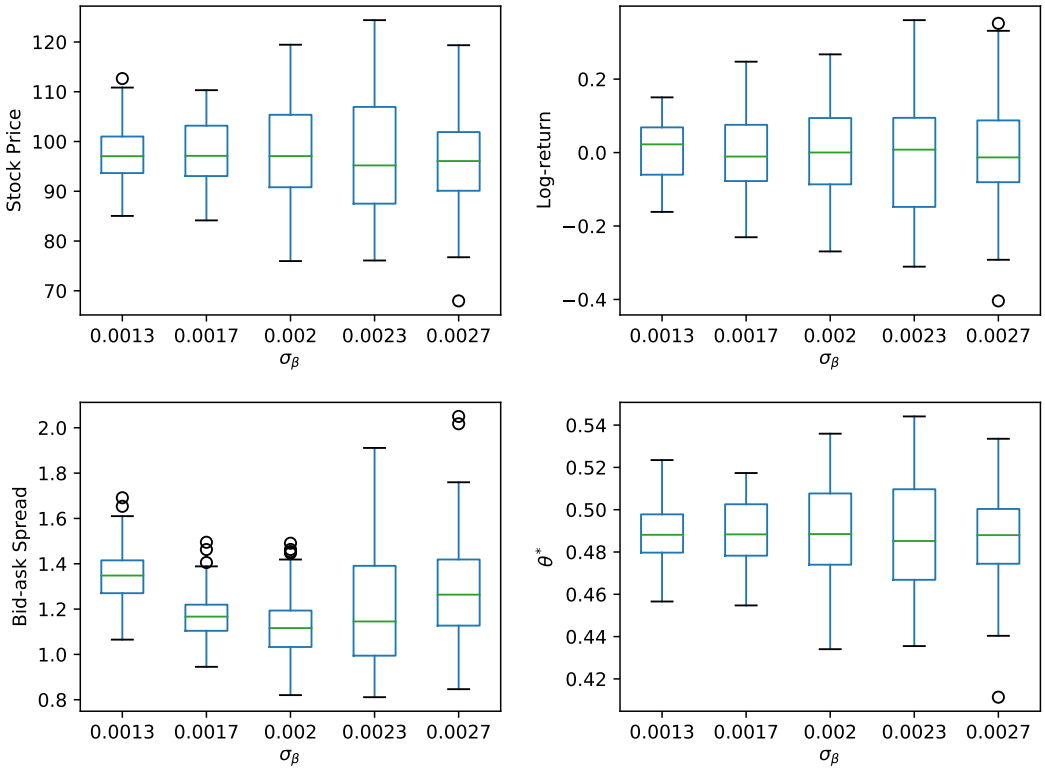


Figure 3.19: The box plots of statistics $\{\tilde{m}_p, \tilde{m}_r, \tilde{m}_d, \tilde{m}_{\theta^*}\}$ against the parameter σ_β.

between the set of statistics $\{\tilde{m}_p, \tilde{m}_r, \tilde{m}_{\theta^*}, \tilde{m}_\theta\}$ and the parameter σ_β. However, a similar shape of curve is observed in the plot of the mean of average stock price $\tilde{m}_p$ against σ_β (top-left panel)and the mean of average terminal optimal risky asset proportion among agents $\tilde{m}_{\theta^*}$ against σ_β (bottom-right panel). It indicates the statistic $\tilde{m}_{\theta^*}$ determines the movement of the mean of average stock price $\tilde{m}_p$. On the other hand, the mean of average bid-ask spread $\tilde{m}_d$ exhibits a U-shape of curve against parameter σ_β, whose the minimum is obtained at $\sigma_\beta = 0.002$.

The sensitivity analysis of parameter σ_β suggests that it can affect the variation of independent model outputs, such as m_p, m_r and θ^*. However there is no significant effect on the median of independent model outputs. It is worthy to note that the movement of $\tilde{m}_p$ is control by the optimal portfolio strategy $\tilde{\theta^*}$. Moreover, there exists a possible quadratic relationship between the mean of the bid-ask spread and parameter σ_β.

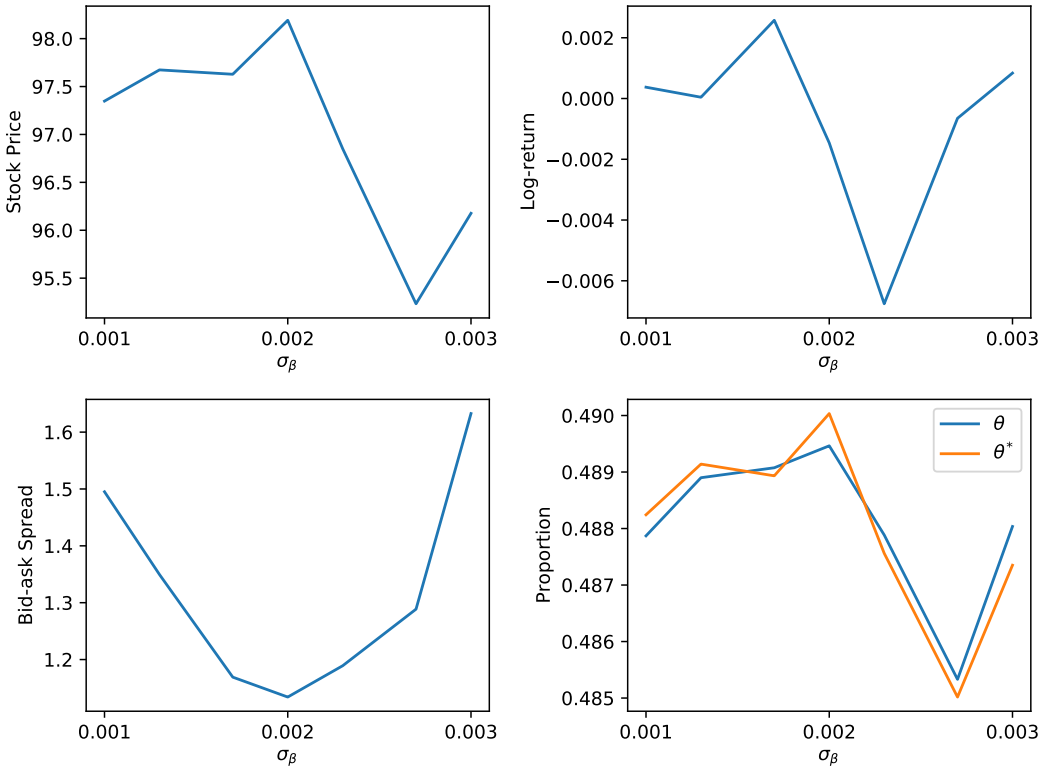


Figure 3.20: The sensitivity test of statistics $\{\tilde{m}_p, \tilde{m}_r, \tilde{m}_d, \tilde{m}_{\theta^*}, \tilde{m}_\theta\}$ against the parameter σ_β.

3.11.3 Model Outputs against Parameter v_0

At last, we investigate how the parameter v_0 influences the statistic $\tilde{S}$ in (3.68). In fact, the parameter v_0 determines the learning rate of agents' beliefs β_t. The value of v_0 is selected evenly starting from 0.001 to 0.003 with a step size $\Delta = 0.0003$. While the value of μ_β is set as 0.005 and the parameter σ_β equals to 0.001. The other parameters remain the same as stated in Section (3.11).

Figure 3.21 presents the box-plot of each element in a set of desired statistics $\{m_p, m_r, m_d, m_{\theta^*}\}$ against parameter v_0. It can be directly viewed that the mean of the stock price $\{m_p^1, m_p^2, \ldots, m_p^H\}$, bid-ask spread $\{m_d^1, m_d^2, \ldots, m_d^H\}$ and optimal proportion at terminal time $\{m_{\theta^*}^1, m_{\theta^*}^2, \ldots, m_p^H\}$ globally decreases as the value of v_0 increases. Moreover, the deviation of the mean of bid-ask spread $\{m_d^1, m_d^2, \ldots, m_d^H\}$ gradually describes against parameter v_0. There is no significant pattern change of the mean of log-return $\{m_r^1, m_r^2, \ldots, m_r^H\}$ against v_0.

Figure 3.22 shows how the change of parameter v_0 affects the average performance of statistics $S = \{m_p, m_r, m_d, m_{\theta^*}, m_\theta\}$. In the top-left panel, the $\tilde{m}_p$ show a negative

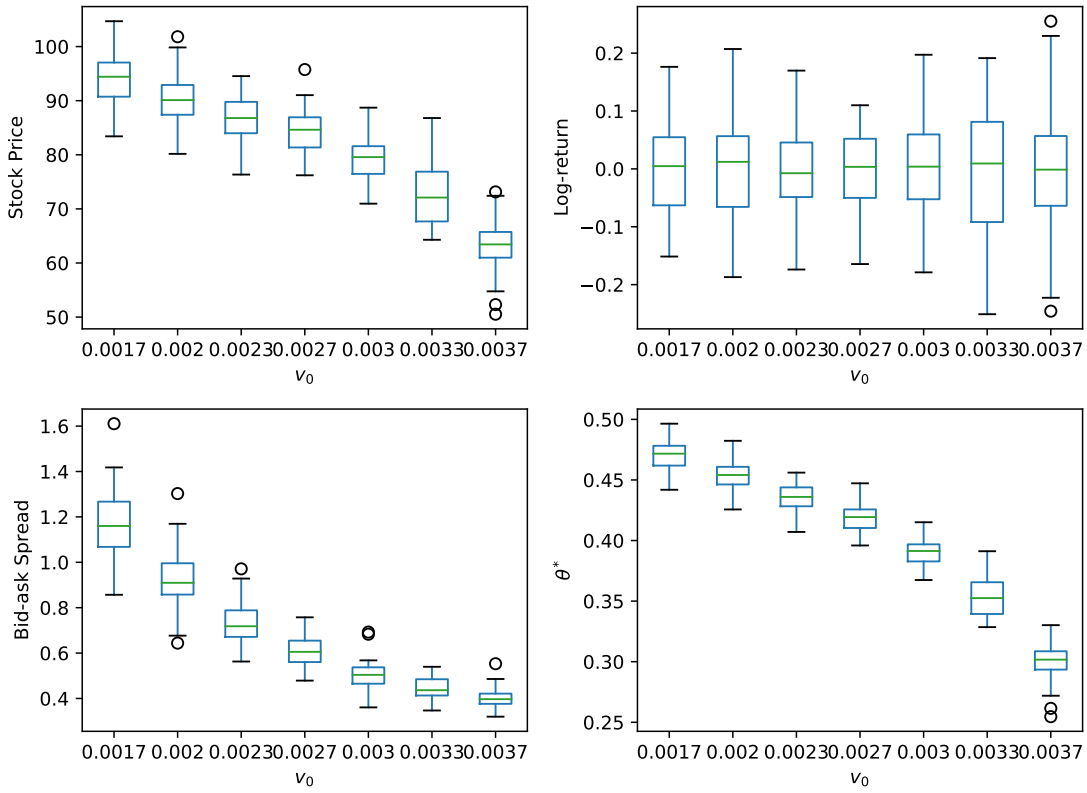


Figure 3.21: The box plots of statistics $\{\tilde{m}_p, \tilde{m}_r, \tilde{m}_d, \tilde{m}_{\theta^*}\}$ against the parameter v_0.

relationship against parameter v_0 and the decreasing accelerates when $v_0 \leq 0.0027$. By contrast, the $\tilde{m}_d$ exhibits a negative relationship against parameter v_0 with a slow decay. In the bottom-right panel, the $\tilde{m}_{\theta^*}$ overlaps with $\tilde{m}_\theta$ showing a negative relationship with parameter v_0 with a accelerating decreasing. There is no obvious influence of average performance of mean log-return $\tilde{m}_r$ against parameter v_0.

The sensitivity analysis of parameter v_0 indicates that agents are more likely to obtain their optimal portfolio strategy goal with a higher value of v_0. There is a positive feedback loop concerning parameter v_0 that a higher value of v_0 makes agents learn faster about the price drift μ_t. An observed low drift of price μ_t leads to a decreasing belief β_t and a lower optimal proportion of risky asset θ_t^*. When $\theta_t > \theta_t^*$, agents will move in the same way to sell the risky asset, which in return causes a lower observed drift of price μ_t. It explains why there is an accelerating decreasing in the plot of $\tilde{m}_p$ against v_0 and $\tilde{m}_{\theta^*}$ against v_0.

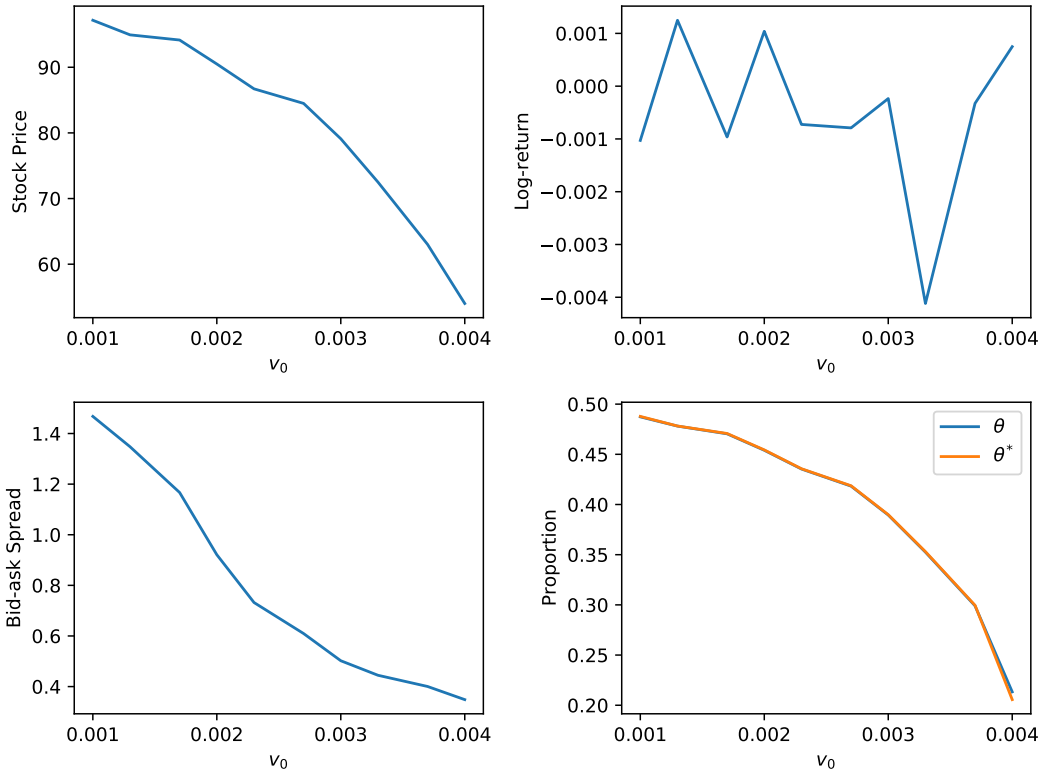


Figure 3.22: The sensitivity test of statistics $\{\tilde{m}_p, \tilde{m}_r, \tilde{m}_d, \tilde{m}_{\theta^*}, \tilde{m}_\theta\}$ against the parameter v_0.

3.12 Summary

This chapter introduced the trending-following agent who uses the Bayesian learning method to track the risky asset return. The idea is derived originally from a portfolio choice problem proposed by Bismuth *et al.* [53] (2019). Therefore, a new ABM was investigated by replacing the noisy agents with the trending-following agents. Finally, we performed a parameter analysis test to identify the vital parameter for determining the statistical properties of the model outputs.

The portfolio choice problem was discussed in two cases by assuming agents have constant relative risk aversion (CRRA) and constant absolute risk aversion (CARA) utility function, respectively. The closed-form solution of the HJB equation can be obtained and applied by agents to either hold the optimal amount of cash M^* investing in a risky asset with a CARA utility function, or maintain an optimal proportion $\theta*$ of investment in a risky asset with a CRRA utility function.

In simulation results, agents are assumed to have a CRRA utility function with $\gamma = 1$, which makes their optimal strategy θ^* proportional to their learning beliefs. We

first tested the case that agents hold fixed beliefs through the simulation and found that the price movement is similar to the ABM with noisy agents that oscillate around a constant dominating by the average optimal risky asset proportion of agents $\widetilde{\theta^*}$. In the case that agents have learning beliefs, a positive feedback loop was observed that the optimal strategy guides agents to reduce their investment in a risky asset when $\theta^* < \theta$. Then a substantial amount of ask orders pull down the price leading to a lower optimal proportion θ^*.

Finally, the parameter sensitivity analysis provides a way to choose a suitable initial range of parameters for the model simulation. Any wrong choice of the initial parameter will cause abnormal simulation results, such as a flat price movement due to no completed transaction and a positive feedback loop in price. The flat price movement can be explained by the reasons that: (1) agents beliefs locate in a small range causing the maximum distance of limit order price to be smaller than the minimum tick size of stock price when μ_β is very tiny; (2) agents beliefs are widely distributed that the best bid and best ask order will never crossover with each other to make a transaction, when μ_β is relatively large. While the positive feedback loop is more likely to be viewed with a higher value of v_0, where agents learn faster about the asset return.

Chapter 4

Calibration and Testing: Genetic Algorithms and Value at Risk

In this chapter, we are going to demonstrate a framework for using a genetic algorithm (GA) to calibrate an Agent-based model (ABM), such as the one in Chapter 2. We aim to find the optimal parameters of an ABM given the historical time series of stock market returns as our input observations. Subsequently, we wish to use simulations from the ABM to predict the distribution of log-returns of observations into the future. First, to demonstrate that our methods work in a controlled environment, we use the simulated data from our model to generate the historical time series and then try to calibrate against it. Next, we use real-world data sets to assess the model in a more realistic setting. In order to test the model, we show a range of results, along with the results of some naive probability distribution fitting approaches to give our comparisons more context. As part of this process, we will also aim to investigate the statistical properties of the observations and simulated data sets to find a suitable metric, which can be used to determine the best calibration and simulation process.

In order to capture the main features of data, we propose that the objective function of a GA should be a linear combination of the log-return moments. We show how this objective function reacts to different input parameters, controlling the volatility of simulated data sets through several parameters. We notice that the complexity and stochasticity of our model do not allow us to find a global optimum. Instead, a high-performance objective function can find a region close to an optimal value, where slight disturbances in the set of input parameters generate only an insignificant difference in

the value of the objective function and the statistical properties of output simulations.

In general, during the calibration process of an ABM, we found it hard to identify why we chose one specification of model parameters rather than another. However, we believe that it can still be useful for predictive purposes, experimentation or better understanding of the underlying dynamics of the real data. Therefore, we can apply the GA calibration result to predict the future empirical distribution of log-returns given a price series input and calculate risk metrics, such as Value at Risk (VaR) and Conditional Value at Risk (CVaR) at different confidence levels, with the idea that this method could be used for risk management. Moreover, we use the real Cryptocurrency data as a training set and empirically observed future distribution as a baseline. We can compare our ABM via a GA calibration approach to other simple statistical distribution models and identify which model can provide the best performance in matching future empirical distribution.

We have contributed to the diverse field of agent-based modelling by identifying an objective with linear combinations of moments provides best fit to the data. We have shown how the GA is robust when it comes to calibration of stochastic outputs. Unfortunately, even with reliable calibrations, we find that ABMs are not significantly better at predicting the risk of investing in assets.

In this chapter, we first review the background of the Genetic Algorithm and introduce a framework to calibrate the ABM introduced in Chapter 2. Then, a calibration experiment is conducted with known inputs. The calibration result is used to predict the future empirical distribution, and two risk metrics, VaR and CVaR, is calculated. Finally, we calibrate an ABM model with Cryptocurrency data as inputs and compare the results with traditional probability distribution fitting approaches.

4.1 An Introduction to Genetic Algorithms

Before moving onto describing how the Genetic Algorithm (GA) works with our model, we provide a brief overview of the method. This is important because certain features (such as the fact no derivatives are required to be calculated as part of the process) are important for a model such as our that generates stochastic outputs. The algorithm itself is inspired by the procedure of natural selection, and it has been widely used

to produce high-quality solutions to optimisation and search problem in a variety of fields. The main functions inside the algorithm are mutation, crossover and selection, which are supposed to mimic those we have observed as part of natural selection in the real world.

Unlike the traditional search or optimisation algorithms, GAs are probabilistic search procedures designed to deal with models with a large set of parameters and no closed form solution for distribution. In [50], Rogers and Von Tessin (2004) proposed an ABM of a financial market and calibrated it using a multi-objective GA. The model evolves a population of Pareto-optimal parameters sets, in which a single candidate can be selected as a final tuned parameter set without requiring any prior explicit weighting of criteria. In [51], a classic model presented by Farmer and Joshi (2002) (see [22]) is calibrated using a Nelder-Mead simplex algorithm and a GA. Both methods can be candidates to fill the gap of parameters estimation in ABMs field and more robust in noisy environment.

To implement a genetic algorithm, we must begin with a population of chromosomes, which can be encoded to represent all possible combinations of parameters within a finite space. We can then evaluate these structures, and allocate reproductive opportunities to each one so that there is a higher chance of chromosomes that give a better solution to the objective function passing their characteristics to the next generation than chromosomes that give poorer solutions. In the setting of an agent based model, where individuals in the model could have many parameters (such as initial number of stocks, cash etc), we believe the genetic algorithm seems appropriate. From this view point, the natural selection process could be seen to be selecting the most appropriate set of agents that can reproduce the input data. We also note that the genetic algorithm is derivative-free, which means it does not rely on derivative information in the classical sense to find optimal solutions. Sometimes the derivative information of the objective function might be unavailable, unreliable or impractical. In specific, the objective function can be non-smooth, time-consuming to evaluate, or in our model having randomised outputs (See Equation in (4.5)). In fact, we tried to follow a calibration procedure, which implements Stochastic Gradient methods in a popular ABM developed by Brock and Hommes (1998) (BH model, see [47] and [48]). Some of our own experiments on this method can be found earlier on in this

chapter, but the results were not very satisfactory. On the one hand, the BH model is analytically tractable involving few parameters and no noise in prices formulation makes the calibration procedure deterministic. On the other hand, our model adds randomness in the operations, such as expectation generation, order placement and behaviour imitation, which gives no closed-form solution for price and the corresponding derivatives.

A typical genetic algorithm consists of the following major steps,

1. initialise a population of solutions
2. select the first generations from the population
3. evaluate the current generation using the objective function and fitness function
4. implement the genetic operators, such as selection, crossover and mutation to produce the next generation
5. repeat the steps (3) - (4) until reaching the maximum number of generations or a predefined fitness level is satisfied.

An important part of applying the genetic algorithm is to encode the range of available inputs to the model as a set of variables that represent the genes or chromosomes in natural selection. The most common way to encode the inputs is to convert them to a binary string (or vector), so that each gene or chromosome is either 'on' or 'off', and the algorithm has been shown to perform best when this binary encoding is used. For our model, in which the inputs are real numbers or integers over an interval, we can easily cover this space by setting up discrete set of binary numbers corresponding to positions within the interval. For example, an input parameter α chosen from the interval $[a, b]$, can be encoded using n genes or chromosomes as

$$\alpha_i = a + i\frac{b-a}{2^n - 1}$$

where the number i is encoded as a binary string s of length n in the algorithm. For example, with $a = 0$, $b = 1$ and $n = 6$, the following set of genes would give

$$s = \boxed{0}\boxed{1}\boxed{0}\boxed{1}\boxed{1}\boxed{1} \quad \rightarrow \quad i = 23 \quad \rightarrow \quad \alpha_{23} = 0.3651.$$

This gives a fixed set of possible inputs at equidistant intervals across the input space.

Figure 4.1 shows how the new generation are produced by implementing the basic GA operators, such as selection, crossover and mutation. For illustration, we start with a current population consisting four strings labelled with different colours. In a selection phase, string 2 is duplicated and substitutes string 3. Selection (or reproduction) is a process that makes more copies of better strings in a new population. Meanwhile, it causes the fact that individuals which encode successful structures have a higher probability to reproduce themselves.

Sequentially, string 1 and string 2 participate in a crossover operation that some portion of these two strings are exchanged with each other to create two new individuals for the successive generations. The objective of a crossover operator is to recombine two strings to get a better one, which together with a mutation operator offset a limitation of the selection phase that no new strings are created.

Here is a simple example shows that two 8 bits binary strings swap their last 5 digital numbers to generate two new strings after a crossover operation.

string 1	011|**01100**	string 1	011|**11001**
string 2	110|**11001**	string 2	110|**01100**
	Before crossover		After crossover

Finally, a mutation operator disturbs genetic information of string 4, which happens at the bit level that each bit may become mutated. The mutation operator aims to add new information in a random way and ultimately help to get rid of a trap at local optima. Moreover, it introduces diversity in the population to avoid a trend of becoming homogeneous after frequent use of selection and crossover operators. An intuitive example explains the importance of mutation operator as we consider the population having four 8 bit strings,

$$01101011 \quad 00111101 \quad 00010110 \quad 01111100.$$

It is worthy to note that all strings have a 0 in the first bit position. If the true optimum solution starting with a 1 in that position, then neither selection nor crossover operator enable us to turn 0 into 1. The inclusion of mutation solves the concern by allocating probability of turning 0 to 1.

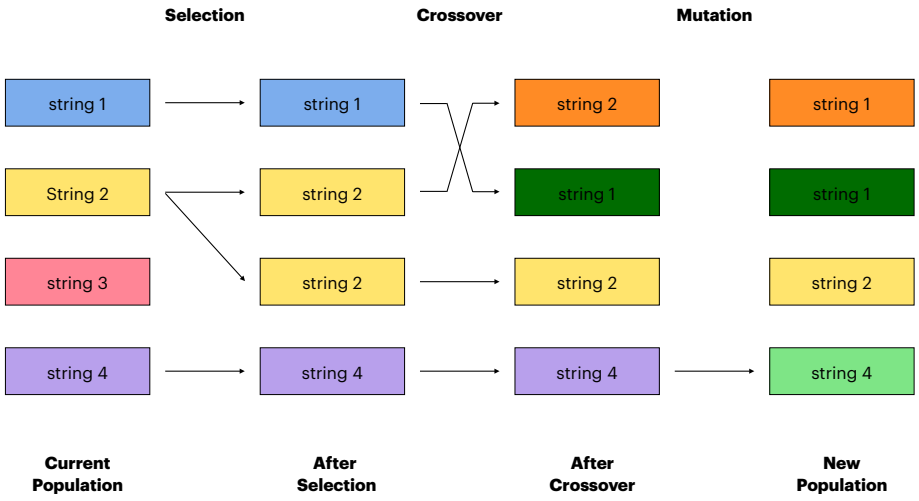


Figure 4.1: The new population are generated by experiencing the basic GA operations, such as selection, crossover and mutation

There are several existing Python libraries for implement a GA in our model. We choose one called PyGaD developed Gad and Fawzy (2021) (see [66]), which is designed specifically for the single-objective optimisation and is user-friendly for customising the fitness function, population, gene value range, gene data type, selection, crossover and mutation.

4.2 Calibrating Parameters: Known Inputs and Outputs

In this section, we firstly present a pseudocode of an GA calibration, which can give us a better understanding of following simulation results.

We move on to show how using ABMs can predict the future distribution of stock returns given some observations that can be used as a training data set. Then, two main problems initially stand out: How can we find the best parameters of an ABM to fit the observations? And, in order to fit parameters, which objective function should we decide on? It seems appropriate then that we should apply the calibration process on a known input, so we choose to try and calibrate against three artificial stock paths that are generated using the ABM with known parameters. Then we pretend that those parameters are unknown, passing the raw data to the GA so that it can find

Algorithm 3 The pseudocode of an GA calibration

Require: r^{obs}, $\mathcal{M}$ (see Equation (4.4)), $N^{generation} \in \mathbb{Z}^+$, $N^{iter} \in \mathbb{Z}^+$, $tol > 0$
  select the first generations from the population $\mathcal{M}^0 = \{\underline{\Theta}_i^0\}_{i=1}^{N^{generation}} \subset \mathcal{M}$
  $N^{iter} \leftarrow 0$
  $n \leftarrow 0$
  $\Delta \leftarrow 1$
  while $n < N^{oter}$ and $\Delta < tol$ **do**
    set the generations $\mathcal{M}^n$ as input parameters in the ABM in Chapter 2
    follow the **Algorithm 1** producing corresponding simulation outputs $r^{sim}(\underline{\Theta}_i^n)$
    evaluate generations using the objective function (4.5) and fitness function (4.6)
    $\Delta \leftarrow \min_{\underline{\Theta}_i^n \in \mathcal{M}^n} f\left(r^{obs}, r^{sim}(\underline{\Theta}_i^n)\right)$ (see Equation (4.5))
    $n \leftarrow n + 1$
    produce new genration $\mathcal{M}^n$ using genetic operators
  end while
  produce the calibrated generations $\mathcal{M}^n$

the best parameters fitting the training data. Obviously, we need to give some way to decide how close our simulated outputs are to matching the real data, so we must explore what a suitable objective function should look like, in particular what function will be most sensitive to the statistical features of the initial data.

In the following part, we will calibrate the parameters of the ABM introduced in Chapter 2. The contents including agent's future expectation generation, orders placement, the communication network and the mechanism of imitation will be directly applied. Let us have a quick recalling of these contents in Chapter 2. The noisy agent i randomly assign generates his expectation about price return with the following equation,

$$\hat{r}_t^i = \sigma_t^i \epsilon_t, \tag{4.1}$$

where $\epsilon_t \sim N(0,1)$ is a standard normal variable. A positive agent-specific volatility σ_t^i has the following formula,

$$\sigma_t^i = A\left(\sigma_0^i + (1-w)\frac{l_i^{in}}{N^{agent}}\right), \tag{4.2}$$

where l_i^{in} is the number of incoming link of the agent. We slight modify the initial money C_0 give to each noisy agent as

$$C_0 = a_0 \cdot S_0 + 1000 \cdot b, \tag{4.3}$$

where S_0 is the initial number of shares given to each agent and the coefficient a_0 is a constant.

Then, let $\mathcal{M}$ be a feasible parameter space defined as follows,

$$\mathcal{M} = \{\underline{\Theta} = (A, w, b) \in R^3, A > 0, w \in [0, 1], b \in [0, 100]\} \tag{4.4}$$

where A scales the volatility of agent's expected returns, w quantifies the imitation effect among agents and b determines the money supply in the system.

In general, the GA is a search process referring to the survival-of-the-fittest principle of nature, which makes it suitable for solving maximisation problem. Let $\{r_t^{obs}\}$ and $\{r_t^{sim}\}$ represent the log-returns of observed and simulated stock paths and their i-th moments are denoted by m_i^{obs} and m_i^{sim} respectively. Then the objective function minimising the linear combinations of moment difference between observed and simulated log-returns is considered,

$$\underset{\underline{\Theta}\in\mathcal{M}}{min} f\left(r^{obs}, r^{sim}(\underline{\Theta})\right) = \underset{\underline{\Theta}\in\mathcal{M}}{min} \sum_{i=1}^{4} a_i \left(m_i^{obs}(r^{obs}) - m_i^{sim}(r^{sim}(\underline{\Theta}))\right)^2 \tag{4.5}$$

where $a_i, \quad i = 1, \ldots, 4$ are weights of each moment. The fitness function is obtained with a easy transformation,

$$\mathcal{F}(r^{obs}, r^{sim}) = \frac{1}{\underset{\underline{\Theta}\in\mathcal{M}}{min} f(r^{obs}, r^{sim})} \tag{4.6}$$

which converts a minimisation problem to an equivalent maximisation problem. In a GA, individuals in the population are allocated with reproductive traits using fitness, which means that individuals with higher fitness value are more likely to be selected as candidates for further operations. If we are to use this objective function, we still need to decide what the weights should be attached to each moment. To decide on this, we next look through some plots of the how the different moments are affected by the input parameters, so that we may choose to assign larger weights to those relatively unaffected, and lower weights to those showing high sensitivity, so that the objective matches as many features found in the moments as possible.

Figure 4.2 displays first four moments of log-returns generated by agent-based model when using different combination of input parameters w and A. We can quickly spot that the mean and skewness of log-returns do not exhibit any distinct features over the parameter space, and in fact the values of these two moments tend to fluctuate around 0. While the standard deviation shows a gradual change along the diagonal direction and the kurtosis has a significant difference in the bottom left area with

higher values. Therefore, we should allocate a high weight of standard deviation and relatively small weights of other moments to keep the feature exhibiting by the parameters w and A.

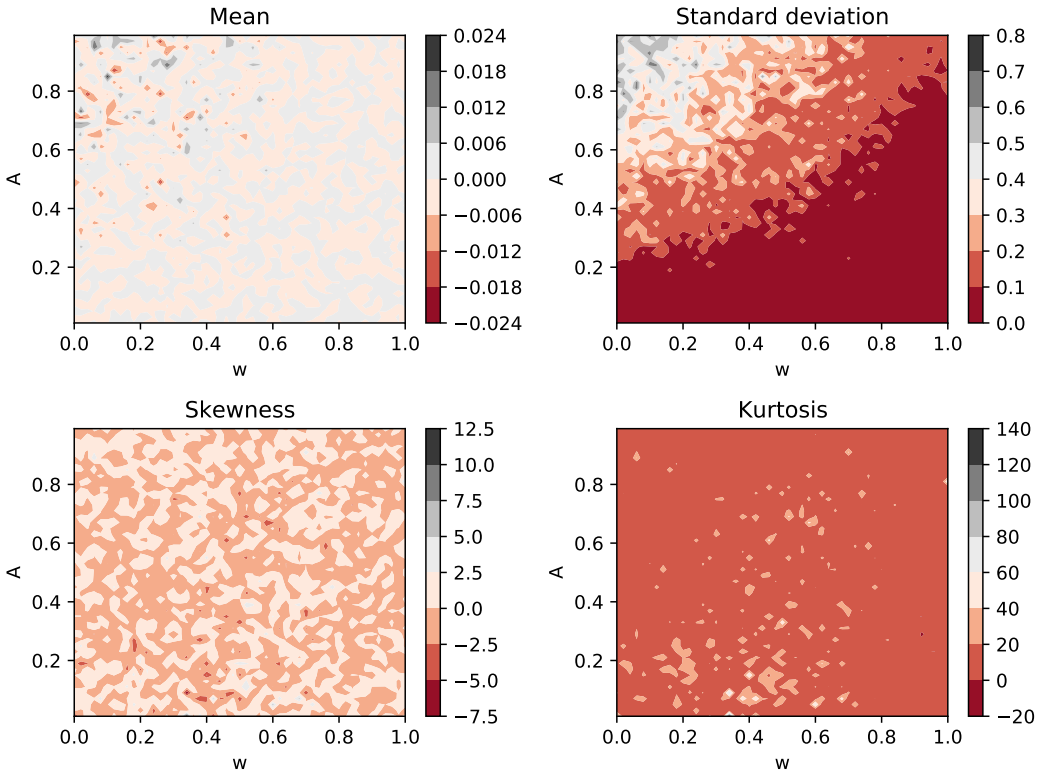


Figure 4.2: The first four moments of log-returns generated by the agent-based model against both parameters w and A.

In practice, we set the weights of each moment as $a_1 = 1, a_2 = 100, a_3 = 0.01, a_4 = 0.01$. The observations we choose are three artificial stock paths generated by agent-based models with parameters $(w = 0.5, A = 0.1)$, $(w = 0.8, A = 0.15)$ and $(w = 0.2, A = 0.02)$ respectively. The resulting value of objective function against parameters w and A given these three observations is shown in Figure 4.3. The results are presented as a heat map, so we can see that the values in the top left are very high, but vary less towards the bottom right. Comparing this with Figure 4.2, we see that the choice of weights result in the second moment (variance) dominating proceedings. The difference between the graphs is barely noticeable to the naked eye, suggesting that the properties of these inputs are broadly similar. The location of the possible parameters are marked on each graph as a cross. At those initial parameter inputs, we calculate the value of the objective function as $f(w = 0.5, A = 0.1) = 2.41$,

$f(w = 0.8, A = 0.15) = 0.33$, and $f(w = 0.2, A = 0.02) = 1.52$. All these values are relatively close to 0 comparing with the region in the top left in Figure 4.2, which indicates a good calibrating performance. Besides, the lowest value is obtained with parameters $(w = 0.8, A = 0.15)$, whose kurtosis is negligible around that area.

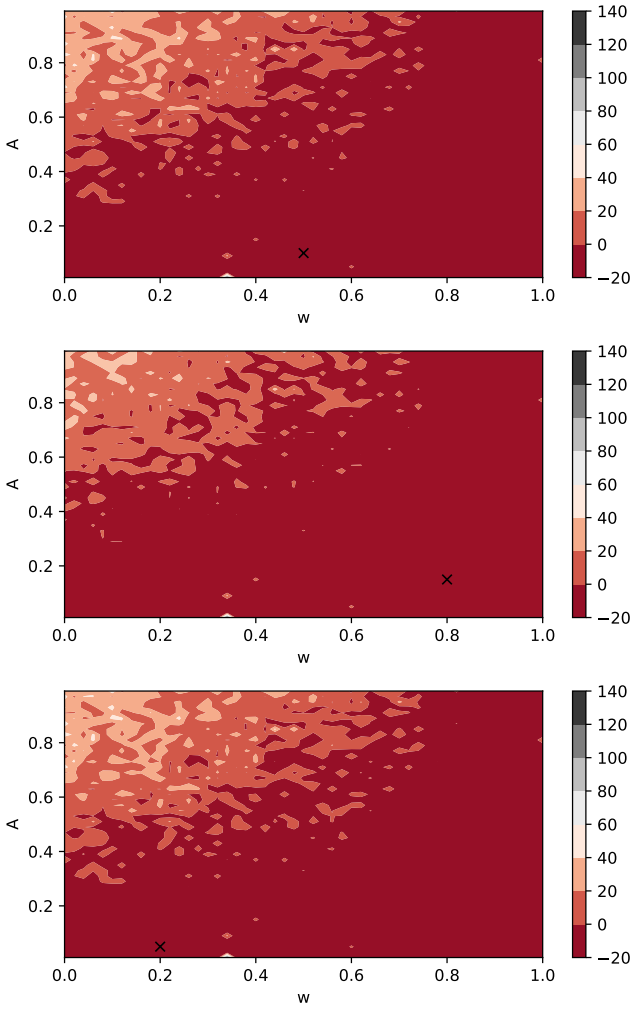


Figure 4.3: The value of the objective function against parameters w and A. The parameters used to generated three artificial observations are marked as cross and their corresponding value of the objective function are calculated as $f(w = 0.5, A = 0.1) = 2.41$, $f(w = 0.8, A = 0.15) = 0.33$, and $f(w = 0.2, A = 0.02) = 1.52$.

Due to the nature of an agent-based model, we unable to guarantee the continuity of the objective function with respect to parameters w and A. However, as mentioned before, we hope to still use it to find a rough area, in which the simulated stock path is statistically closer to the observations than it would be in other areas. We are happy that this choice of objective function and associated weights will provide good enough results for our model.

4.3 Risk Management

Risk management is a procedure for shaping a loss distribution. In this section, we introduce two popular functions value-at-risk (VaR) and conditional value-at-risk (CVaR), which are widely used for measuring financial risk.

Let X be a loss variable of the portfolio with cumulative distribution function $F_X(z) = P\{X \leq z\}$. The VaR of X with confidence level $\alpha \in [0, 1]$ is

$$\mathrm{VaR}_\alpha(X) = min\{z|F_X(z) \geq \alpha\}. \tag{4.7}$$

Essentially, $\mathrm{VaR}_\alpha(X)$ is a lower α-percentile of the random variable X. It is commonly used in finance regulations, such as Basel I and Basel II, in which VaR deviations are measured as the width of the daily loss distribution of a portfolio.

CVaR is an alternative percentile measure of risk (see [67]). For continuously distributed random variables, $\mathrm{CVaR}_\alpha(X)$ equals the conditional expectation of X subject to $X \geq \mathrm{VaR}_\alpha(X)$. The CVaR of X with confidence level $\alpha \in [0, 1]$ is defined as

$$\mathrm{CVaR}_\alpha(X) = \int_{-\infty}^{\infty} z dF_X^\alpha(z), \tag{4.8}$$

where

$$F_X^\alpha(z) = \begin{cases} 0, & \text{when } z < VaR_\alpha(X), \\ \frac{F_X(z)-\alpha}{1-\alpha}, & \text{when } z \geq VaR_\alpha(X). \end{cases} \tag{4.9}$$

4.3.1 VaR and CVaR for the Normal Distribution

Let us assume that $X \sim N(\mu, \sigma^2)$ is a normal random variable and $Z \sim N(0, 1)$ is a standard normal random variable. Then the VaR of X with confidence level $\alpha \in [0, 1]$ over the next one period is,

$$\mathrm{VaR}_\alpha(X) = \{z|P(X > z) = \alpha\}. \tag{4.10}$$

Since

$$\begin{aligned} P(X < z) &= P(\mu + \sigma Z \leq z) \\ &= P(Z \leq \frac{z-\mu}{\sigma}) \\ &= \Phi(\frac{z-\mu}{\sigma}) = 1 - \alpha, \end{aligned}$$

we can deduce that,

$$\mathrm{VaR}_\alpha(X) = \Phi^{-1}(1-\alpha)\sigma + \mu, \tag{4.11}$$

where $\Phi^{-1}(\cdot)$ is the inverse x standard normal distribution.

The CVaR of X over the next one period can be expressed in the term of VaR,

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(X)du. \tag{4.12}$$

Combining with Equation (4.11), we can rewrite $\mathrm{CVaR}_\alpha(X)$ as,

$$\begin{aligned}\mathrm{CVaR}_\alpha(X) &= \frac{1}{1-\alpha}\int_\alpha^1 (\Phi^{-1}(1-u)\sigma + \mu)du \\ &= \frac{1}{1-\alpha}\int_\alpha^1 \Phi^{-1}(1-u)\sigma du + \mu.\end{aligned}$$

Let $y = \Phi^{-1}(u)$ and $du = \phi(y)dy$, we can have

$$\begin{aligned}\mathrm{CVaR}_\alpha(X) &= \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^\infty -y\sigma\phi(y)dy + \mu \\ &= \mu - \sigma\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}\end{aligned}$$

where $\phi(\cdot)$ is the standard normal density function.

4.3.2 VaR and CVaR for the Student t-Distribution

Let us assume that $X = \mu + \sigma T$ and T is a standardised t-distribution random variable with v degrees of freedom. Then the VaR of X with confidence level $\alpha \in [0,1]$ over the next one period is,

$$\begin{aligned}\mathrm{VaR}_{v,\alpha}(X) &= \{z|P(X > z) = \alpha\} \\ &= \sqrt{v^{-1}(v-2)}t_v^{-1}(1-\alpha)\sigma - \mu,\end{aligned}$$

where μ and σ are the location and scale parameter of the loss variable X and $t_v^{-1}(\cdot)$ is quantile function of the t distribution .

The CVaR of X with confidence level α over the next one period can be calculated as,

$$\mathrm{CVaR}_{v,\alpha}(X) = \mu - \sigma\frac{v + (t^{-1}(\alpha))^2}{v-1}\frac{\tau(t^{-1}(\alpha))}{\alpha}, \tag{4.13}$$

where $\tau(x)$ is the standard density function of the t-distribution with v degrees of freedom.

4.3.3 VaR and CVaR for the GA Calibration Log-Return

Let $\{\Theta_i\}_{i=1}^{N}$ be defined as the set of parameters from the last iteration of a GA calibration. Then, each of those parameters are used as the input parameters of a ABM model to further produce 10 independent sample paths in each case, making $10 \times N$ in total. We take the last log-return of all these $10 \times N$ sample paths and sort them in an ascending order, which are denoted as $\{r_i\}_{i=1}^{10\times N}$.

Let us assume that X is a lose variable drawn from the GA calibration log-return $\{r_i\}_{i=1}^{10\times N}$. Then the VaR of X with confidence level $\alpha \in [0, 1]$ over the next one period can be simply approximated as,

$$\begin{aligned}\mathrm{VaR}_{GA,\alpha}(X) &= \{z|P(X > z) = \alpha\} \\ &\approx r_{i=\lfloor 10\times N\times(1-\alpha)\rfloor},\end{aligned}$$

where $\lfloor\cdot\rfloor$ is a floor function.

The CVaR of X with confidence level α over the next one period can be calculated as,

$$\begin{aligned}\mathrm{CVaR}_{GA,\alpha}(X) &= \frac{1}{1-\alpha}\int_0^{1-\alpha} \mathrm{VaR}_{GA,u}(X)du \\ &= \frac{1}{1-\alpha}\int_0^{1-\alpha} Q_(u)du \\ &\approx \frac{1}{1-\alpha}\sum_{i=1}^{\lfloor 10\times N\times(1-\alpha)\rfloor}\frac{r_i}{10\times N} \\ &\approx \frac{\sum_{i=1}^{\lfloor 10\times N\times(1-\alpha)\rfloor} r_i}{\lfloor 10\times N\times(1-\alpha)\rfloor}.\end{aligned}$$

4.4 Results

In the results section, we will try to demonstrate the relative performance of the calibration method by first evaluating how it performs on a simulated set of input

data, before moving on to see how it can perform when tested on real data.

4.4.1 Calibrating against Simulated Data

To show how the GA can find the optimal parameters of agent-based model given observations, we use three artificial stock paths rather than real world data, for which we know the actual value of parameters w and A. Each stock path is generated from a single simulation of the ABM, and has 200 periods (data points). For each period, it consists of 300 intra-day periods. The initial stock price is set as 100 and 100 agents are involved in the market with a portfolio containing 100 shares and 10000 cash. Finally the only other parameters in the model are A and ω which change with the simulations.

Figure 4.4 demonstrates how the parameter populations change over generations using the GA. On the left-hand side, we plot the objective function over the possible state space, with blue points indicating the random initial parameter space's location and the green triangles indicating the location of the final population. We see in all three cases from top to bottom, the initial populations (blue points) are widely spread in the feasible space, whilst the last populations (green triangles) gather around the actual value of parameters (black cross). Also on those figures, we present the best or optimal solution with a red star. As it turns out, we find that the optimal solution is not the closest one to the actual values in each of the three scenarios. Next, on the right-hand side, we plot fitness values as stated in (4.6) for each generation. We see that in all cases, the value of fitness increases over the first 5 or 6 generations. For the top and middle graphs, we see that increase continue up to the 10th iteration and the algorithm can gradually generate a high-quality solution pool to optimise the objective function. For the bottom graph, we see that the fitness value will oscillate past the 5th iteration. This is caused by the fact we have stochastic outputs from the model, meaning that even a correct parameter choice can give a lower fitness value. This can be seen because the optimal value (red star) is often not found close to the real value. We consider the fact that the parameter space is reduced to a small region around the real value providing a good result, and that this is in stark contrast to our attempts to fit the model with gradient-based methods, which failed to converge at all.

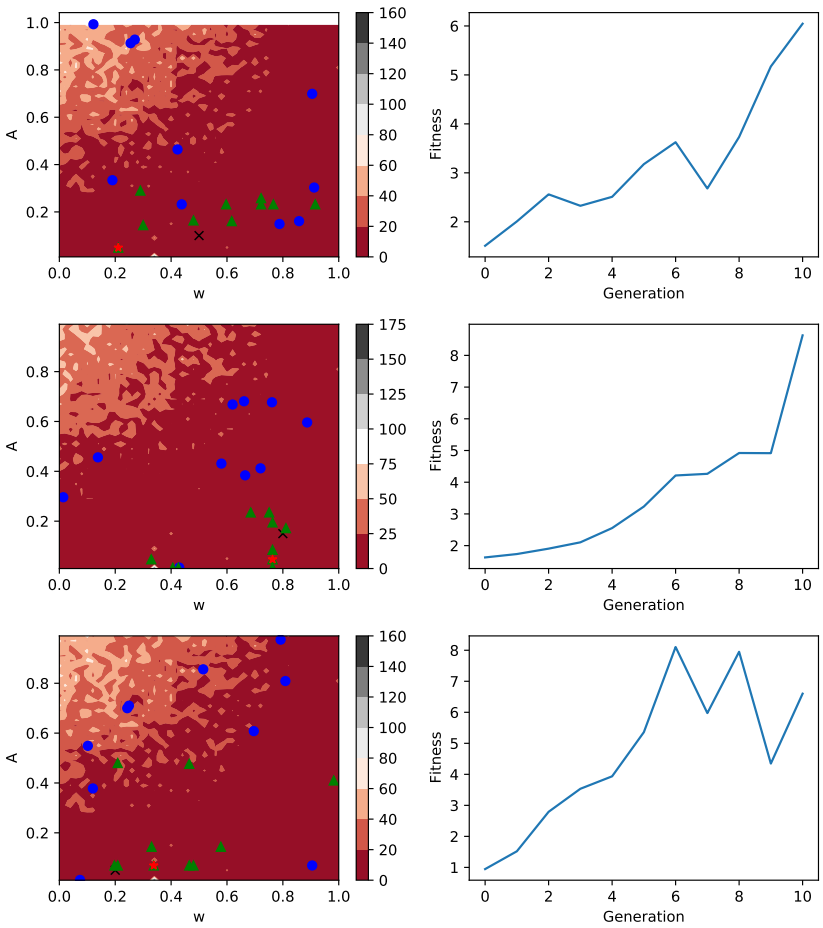


Figure 4.4: Left panels: The evolution of parameter over generations using the GA. The artificial observations are generated by agent-based models with parameters $(w = 0.5, A = 0.1)$, $(w = 0.8, A = 0.15)$ and $(w = 0.2, A = 0.02)$ respectively. The blue points are the initial parent parameters and the green triangles represent the last generation parameters. Red star and black cross stand for the actual value and optimal value of parameters respectively. Right Panels: The corresponding sum of fitness values in each generation.

Following on from the initial results earlier on in this chapter, where the objective function was tested on simulated input data, we move on to develop some metrics by which we can adequately evaluate the calibration process. Two risk metrics VaR and CVaR are calculated via four different methods. For the first two methods, we assume the log return of observations following two popular distributions: normal distribution and t distribution. We can easily get the maximum likelihood estimators of these two distributions given observations as input data. Then the corresponding VaR and CVaR can be directly obtained by formulas introduced in Section 4.3.1 and Section 4.3.2. The third method relies on the fact that observations are artificially simulated

by ABMs, whose parameters we have already known, and we can repeatedly run the model at a certain period as many times as we want to obtain an empirical future distribution of log-returns. For the fourth method, we implement a GA to calibrate the observations and use its population in the last generation as input parameters of ABMs to generate an empirical future distribution of log returns. The equations for calculating VaR and CVaR are introduced in Section 4.3.3.

In the experiments, we assume a portfolio investing 10 000 into a stock, whose path is artificially simulated by the ABM we introduced in Chapter 2. For the ABM simulation path, we only vary the value of w and keep the other parameters the same as stated at the beginning of this section.

In Figure 4.5, we calculate the one-day VaR of the portfolio at different confidence levels under six different scenarios. Considering the observations produced by the ABM simulation, we can assume that the VaRs evaluated by the ABM method reflect the actual empirical future distribution. We can determine whether the other methods can provide a convincing risk assessment by comparing them with the ABM method.

In the case $w = 0$, none of the methods can match the ABM results well and we see that the t-distribution and the normal distribution method both overestimate the VaRs. In comparison, the GA method is heavily affected by the extremums at the confidence level less than 0.03, where the risk assessment is nearly triple of the ABM value. In the case $w = 0.1$, the GA method results become stable and underestimate VaRs compared with the ABM method. When we look at the case $w = 0.3$, the t-distribution and GA method show a good performance in fitting the value of VaR, but the normal-distribution keeps overestimating the values. In the cases that $w = 0.5$, $w = 0.8$ and $w = 1$, the GA method can not properly predict the future return distribution of the observations, whose values deviate far from the ABM values. By contrast, the t-distribution and normal distribution methods can better evaluate the value of VaR as w increases. Especially in the case $w = 1$, their evaluations show no significant difference to the VaRs of the ABM method.

In the view of overall situations, parameter w determines the kurtosis of the simulated sample path, whose value varies from 10.017 to 0.366 indicating more extreme values existing in log-return distribution. (see Table 4.1). Therefore, there is no suitable method to properly evaluate VaRs when the value of w is too small. Besides, the

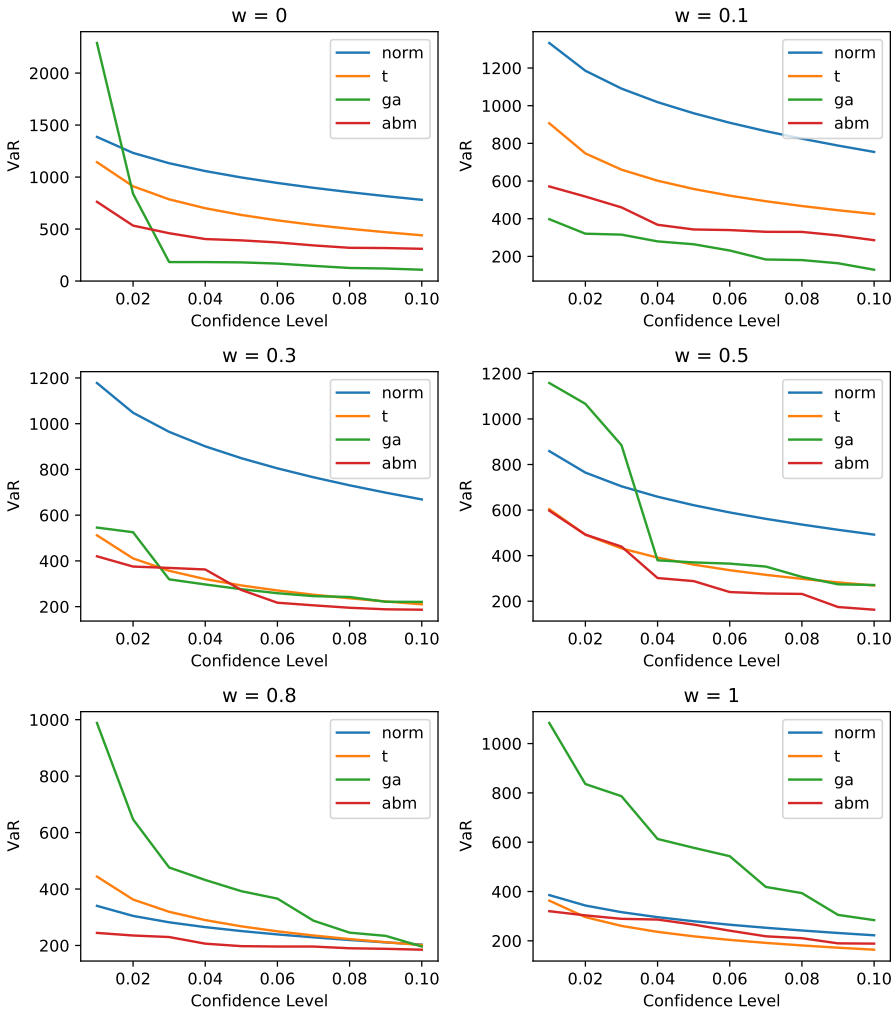


Figure 4.5: One-day VaR calculation given 10000 initial investment: The results are calculated by assuming the log return of stock price following standard distributions, such as the normal and t distribution or the empirical distributions generated via the ABM and GA approach.

t-distribution method should the optimal way to calculate VaRs. the t-distribution method usually gives an optimal assessment. Furthermore, the performance of GA method is relatively poor, which is heavily affected outliers in predicting future return distributions. Therefore we might conclude that the GA method can not sufficiently capture the feature of observations, whose parameters only partially calibrate the observation in volatility.

In contrast, CVaR is a coherent risk measure and is continuous with respect to the confidence level. Figure 4.6 displays the one-day CVaR of the portfolio at different confidence levels under six different scenarios. For comparison, we treat the results via ABM method as the baseline and use it to testify whether the other methods can

provide a convincing risk assessment.

In the cases $w = 0$ and $w = 0.1$, we notice again that none of the methods can match the ABM results well, with the t-distribution and normal distribution overestimating CVaRs, and the GA method gives an unstable evaluation. When we look at the case $w = 0.3$, both t-distribution and GA method can fit the CVaRs well, while the normal distribution method keeps overestimating the risk level. Similar to the situation we see in the VaR calculation, the GA method can not properly match the future return distribution of the observations under the cases that $w = 0.5$, $w = 0.8$ and $w = 1$, and it keeps exaggeratedly evaluating CVaRs. Moreover, the t-distribution method shows a good fitting in the cases that the $w = 0.5$ and $w = 0.8$, which is more accurate than the normal distribution method.

In general, the estimation accuracy of CVaR is heavily influenced by tail modelling. The CVaR value may be misleading if not enough information about the tail is provided. The GA method exhibits the poorest performance in matching the CVaRs, whose simulation returns are prone to the distribution's fatter tails. Apart from the cases that $w = 0$ and $w = 0.1$, the t-distribution method is always the best choice for evaluating CVaRs since it has higher kurtosis than the normal distribution method given the same observations.

Finally, we summarise the results across a range of metrics. We perform the two-sample tests (or homogeneity test), such as Kolmogorov-Smirnov test (KS), Anderson-Darling test (AD) and Cramér-von Mises test (CM) (see [68], [69] and [70]), to judge whether the return distribution between simulations and observations are the same or not.

We produce a set of observations consisting of six log return series using a ABM, in which parameter w varies in a range $[0, 1]$ and the others keep the same as stated at the beginning of this section. For each observed series, we replicate it via three methods. We use the probability distribution fitting approach that the log return of observations is assumed to follow the normal distribution and t distribution, respectively. The maximum likelihood estimators are set as probability parameter inputs to generate random log returns with these two distributions. At the same time, the GA method inputs its optimal solution into the ABM to produce a simulated log return series. In total, there are 18 groups of sample paths performed with two sample tests.

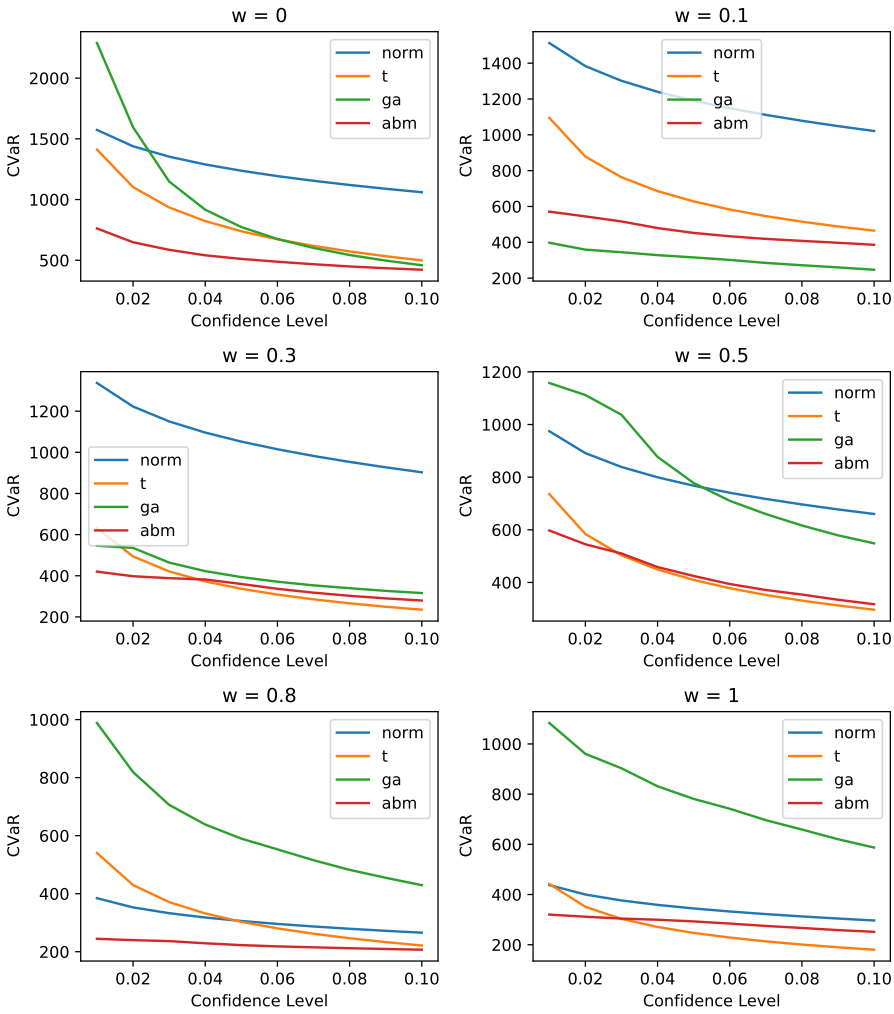


Figure 4.6: One-day CVaR calculation given 10 000 initial investment: The results are calculated by assuming the log return of stock price follows distributions, such as the normal and t distribution or the empirical distribution generated via a combination of ABM and GA approach.

In Table 4.1, the information, like moments of observations, test statistics and the corresponding p-values, are presented. In the case $w = 0$, the log return of observations has a right-skewed and leptokurtosis distribution, skewness (sk) = 1.738 and kurtosis (kr) = 10.017. At the 10% confidence, there is a significant difference between the log return of observations and normal distribution simulations in AD ($p = 0.07$) and CM ($p = 0.07$) test. However, for the groups, observation vs t-distribution simulation and observation vs GA simulation, no significant difference exists in KS, AD and CM test at the 10% confidence. In the case $w = 0.1$, the log return of observations has a left-skewed and leptokurtosis distribution, sk = 1.738 and kr = 10.017. Only the AD test ($p = 0.09$) indicates there is a significant difference

between the log return of observations and normal distribution simulations at the 10% confidence. When we look at the case $w = 0.3$, the log return of observations has a right-skewed and leptokurtosis distribution, sk = 1.293 and kr = 5.264. All the tests, KS ($p = 0.04$), AD ($p = 0.04$) and CM ($p = 0.05$), imply that there is a significant difference between the log return of observations and normal distribution simulations at the 10% confidence. For the rest of cases, all the three tests can not indicate there is a significant difference between target groups at the 10% confidence.

In summary, the normal distribution fitting method is not able to reproduce the observations well in the situation that the log return of observations has a leptokurtosis distribution. However, the reproduction of log return conducted via the t-distribution fitting method and the GA method has no significant difference from the corresponding observations in KS, AD, CM two sample test at the 10% confidence. For the performance of three tests, although the AD and CM test are alternatives of the KS test, the KS test gives a different statistical inference from the AD and CM test in the case $w = 0$ and $w = 0.1$, since KS test is more sensitive to deviations near the centre of the distribution than at the tails.

4.4.2 Simulating against Real Data

Here we move on to fitting our model against some real data. Our ABM aims to capture the herding effect we observe in markets and we know it can also reproduce a set of stylised facts emerging in real financial price series. Given the unregulated nature and large number of small investors, the major cryptocurrency markets seem an obvious place to look for a good fit for this model. Authors have attempted to study cryptocurrencies from different perspectives for the last decades since they exhibit beyond-regular performance. In [71], Corbet *et al.* (2019) stated that the most of studies on cryptocurrencies mainly concentrating on the fields, such as market efficiency(see [72], [73] and [74]) and asset pricing bubbles (see [75] and [76]). However, only a few papers shed light on the behaviour of cryptocurrencies. In [77], Corbet et al. investigated the relationship between a set of digital assets and government financial regulation and control tools, such as US Federal Fund interest rate and Quantitative Easing announcements.

Meanwhile, Kristoufek (2013) studied the interaction between Bitcoin and search

queries on Google Trends and Wikipedia (see [78]). Additionally, Bouri *et al.* (2019) presented evidence for a co-explosive behaviour in seven major cryptocurrencies (see [79]). All these fascinating findings could be evidence of the presence of herding or other potential phenomena in major cryptocurrency markets.

Furthermore, Ballis and Drakos (2020) found that investors are more likely to act irrationally and imitator other's ideas in the cryptocurrency market (see [80],). They studied six cryptocurrencies which are: (i) Bitcoin, (ii) Dash, (iii) Ethereum, (iv) Litecoin, (v) Monero and (vi) Ripple. The corresponding results indicate that herding is indeed occurring in these cryptocurrencies and this herding behaviour could be a partial explanation of the co-explosivity phenomena reported in [79].

Since herding behaviour is more likely to be observed in major cryptocurrencies, we choose daily close price of Bitcoin during the period from March 2021 to September 2021 as our observations. During these six month periods, we further divide them into three phases: (1) declining trend (2) increasing trend (3) flat trend, with each phase containing 50 observations. We aim to test the performance of GA calibration in common market trends and compare the accuracy of risk assessment using different evaluation methods.

In Figure 4.7, we present the trajectory of the closing price of Bitcoin and the corresponding quantile-quantile (Q-Q) plot of its log returns in all three phases. In the top panel, the price of Bitcoin has an obvious downward trend, whose average daily log return is -1.03%. In the middle panel, the closing price climbs up with an average daily log return equalling 0.85%. For the bottom panel, the closing price has a mean reversion pattern, and its average log return equals -0.13%. The Q-Q plot against the standard normal distribution indicates that the distribution of the log return in three phrases has a similar shape of density curve and a light tail, whose median varies around 0. We aim to figure out how the GA performs in calibrating real data with different drifts. Furthermore, we introduce a new parameter b (see Equation (4.4)), which determines the initial money C_0 given to each agent in the ABM from chapter 2 rather than a constant. The initial money C_0 is designed to be

$$C_0 = a_0 \cdot S_0 + 1000 \cdot b \tag{4.14}$$

where S_0 is the initial number of shares given to each agent. The coefficient a_0 is a constant, and we set it as $a_0 = 30000$ in the following calibration procedure.

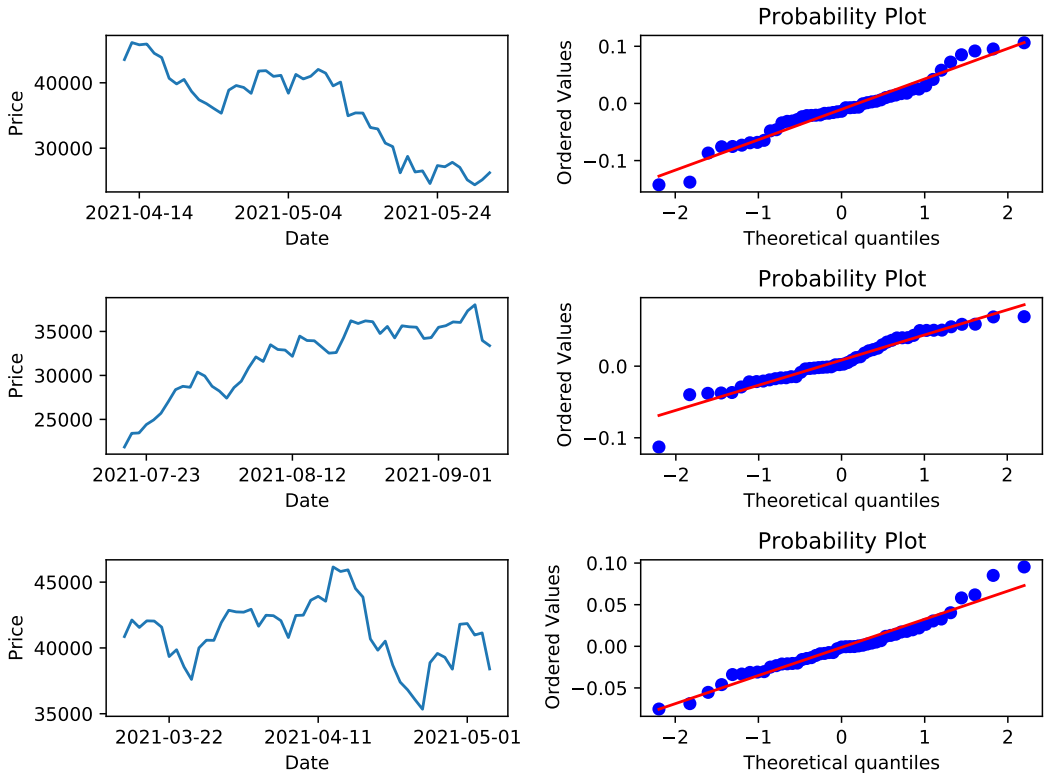


Figure 4.7: Left panels: The closing price of Bitcoin selected from three different periods that the average log return of the price is -1.03%, 0.85% and -0.13%, respectively. Right panels: The corresponding Q-Q plot of log return against the standard normal distribution.

Then a GA is implemented to find the optimal parameters of the ABM in a feasible parameter space $\mathcal{M}$ (see Chapter 2 and Equation (4.4)), which can best fit the observations by maximising the fitness function in Equation (4.6). For the ABM, we set the number of period as 50, within which there are 300 intra-day periods. The initial stock price is the same as the first data point of the observation, and we assume 100 agents are involved in the ABM market with a portfolio containing $S_0 = 100$ shares and $3\,000\,000 + 1\,000 \cdot b$ cash. Meanwhile, the coefficients in the objective function (4.5), which intents to minimise the linear combinations of moment, is set as $a_1 = 1, a_2 = 100, a_3 = 0.01, a_4 = 0.01$.

Figure 4.8 demonstrates how the parameter populations change over generations using the GA to calibrate Bitcoin price. From top to bottom, we sequentially display the result of calibrating Bitcoin price in decline, increasing and flat phases and their optimal or best solution are $(A = 0.368, w = 0.986, b = 9)$, $(A = 0.309, w = 0.972, b = 21)$ and $(A = 0.283, w = 0.995, b = 97)$, respectively. In the left panels, the blue

points indicate the random initial parameter A, and w over the possible state space, while the green triangles represent the last location of the final population. The red star is the optimal or best value of (A, w) found by the GA. As we can see that the initial populations (blue points) are widely spread in the feasible space, while the scope of (A, w) narrows down to a smaller area in the last generation (green triangles) and gathers around the optimal solution (red star). In the right panels, the initial money C_0 given to each agent is presented against generations. In the declining phase, the initial closing price of Bitcoin is over 40 000 and the best solution of parameter b indicates the initial money given to each agent should be 3 009 000. According to the order formation mechanism designed in the ABM, the initial expectation of simulated bid volume is around 38 (see Equation (2.5)), while the expectation of initial ask volume is 50. The stronger power of selling will force the simulation price to move down, by which the ABM can capture a declining trend we observed in real data. Similarly, the initial closing price of Bitcoin is around 22 000 in the increasing phrase, and the corresponding best solution of parameter b implies the initial expectation of simulated bid volume is 69 higher than the expectation of initial ask volume 50. The stronger power of buying guarantee a increasing trend in the simulated path. Finally, in the flat phase, the initial closing price of Bitcoin is around 41,000. In order to capture the mean reversion feature of the real data, we expect to obtain a big value of parameter b from the calibration process, such that we can reach an equilibrium between the buying and selling power. However, the parameter b is an integer variable constrained in a pre-defined range $[0, 100]$. Then the best solution of parameter b is given as 97.

Here we move on to evaluate two risk metrics, VaR and CVaR via four different methods, which are norm, t, GA and historical method. We aim to justify whether the GA calibration combined with an ABM can provide a satisfactory risk prediction in the real world.

The norm, t and GA method have been well introduced in the Section 4.4.1, while the historical method sorts the log return of observations in an ascending order within recent N periods. Let us assume that X is a loss variable drawn from the sorted historical log return $\{r_i\}_{i=1}^{N}$. Then the VaR of X with confidence level $\alpha \in [0, 1]$ over

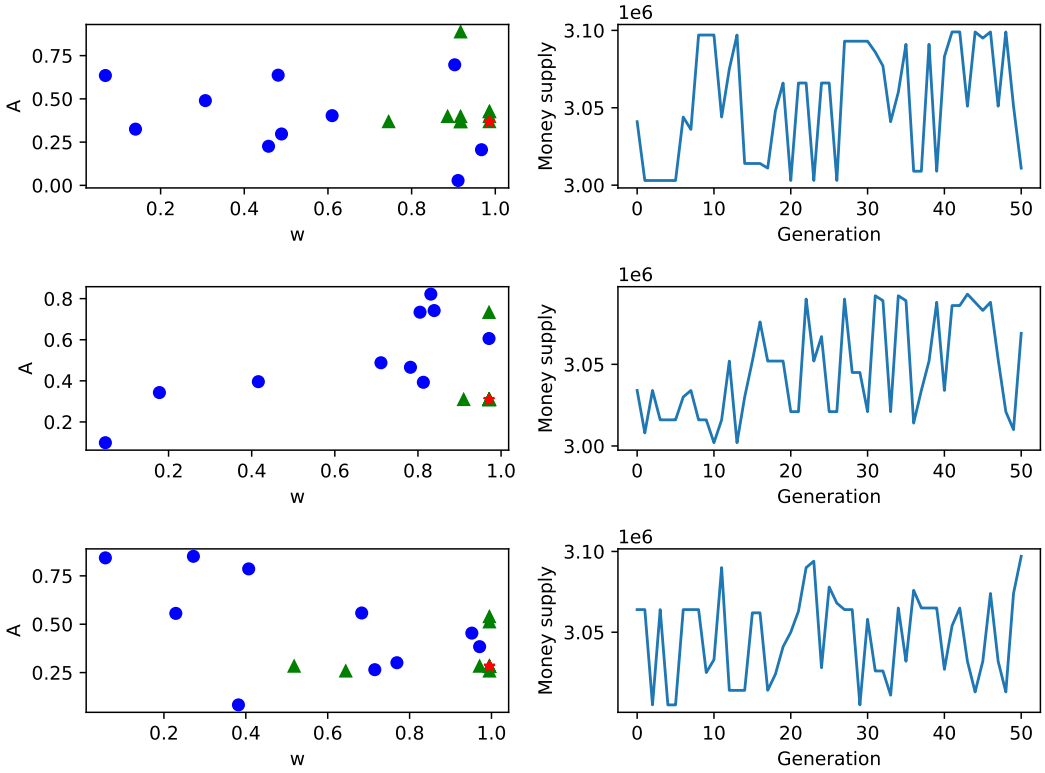


Figure 4.8: The GA calibration result: Left panels: The evolution of parameter (A, w) changes over generations that the blue points stand for the initial populations and the green triangles represent the final populations. The optimal or best solution of the GA calibration is marked as a red star. Right panels: The initial cash C_0 given to each agent changes over generations when we use an ABM to calibrate the real data.

the next one period can be simply approximated as,

$$\begin{aligned} \text{VaR}_{hist,\alpha}(X) &= \{z | P(X > z) = \alpha\} \\ &\approx r_{i=\lfloor 10 \times N \times (1-\alpha) \rfloor}, \end{aligned}$$

where $\lfloor \cdot \rfloor$ is a floor function.

The CVaR of X with confidence level α over the next one period can be calculated as,

$$\begin{aligned} \text{CVaR}_{hist,\alpha}(X) &= \frac{1}{1-\alpha} \int_0^{1-\alpha} \text{VaR}_{hist,u}(X) du \\ &\approx \frac{\sum_{i=1}^{\lfloor 10 \times N \times (1-\alpha) \rfloor} r_i}{\lfloor 10 \times N \times (1-\alpha) \rfloor}. \end{aligned}$$

In Figure 4.9, we calculate the one-day VaR and CVaR of the portfolio investing 10 000 into Bitcoin at different confidence levels in three phases. For comparison, we set the value from the historical method as a baseline (labelled as Bitcoin in the figure), and the performance of the other methods can be roughly judged by how well the curve can match the baseline.

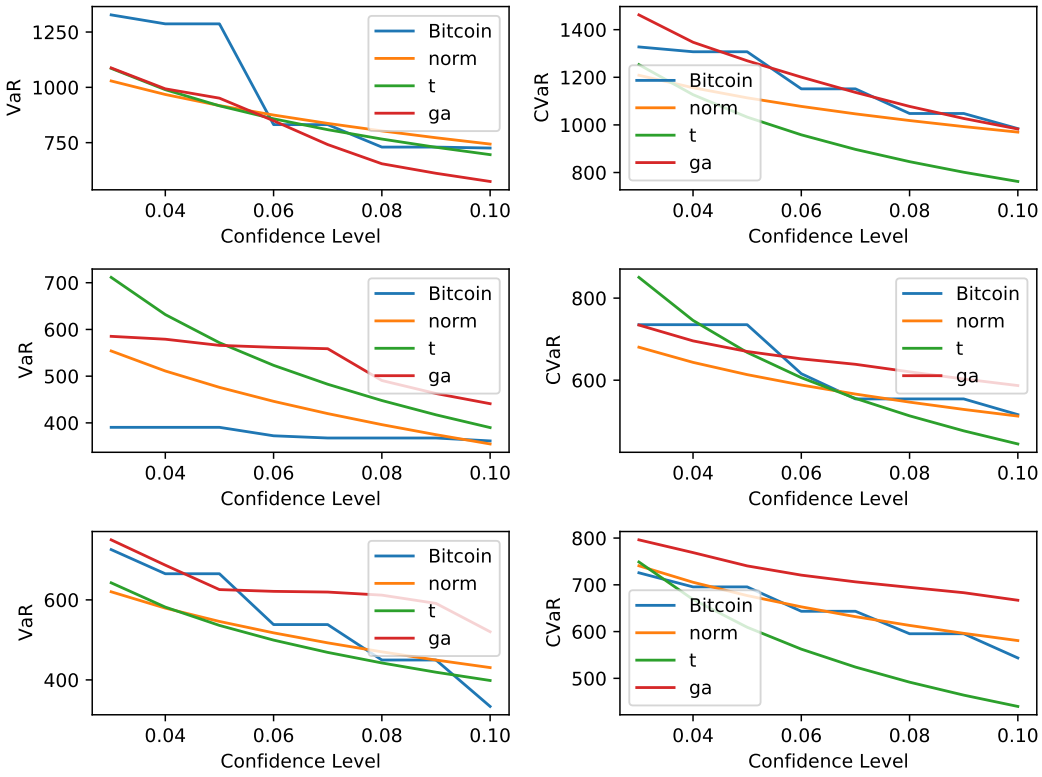


Figure 4.9: One-day VaR (left panels) and CVaR (right panels) evaluation given 10 000 initial investment: From top to bottom, the results are displayed sequentially in declining, increasing, flat phases by assuming the log return of price follows distributions, such as the normal and t distribution or the empirical distribution generated via a combination of ABM and GA approach. For comparison, we set the value from the historical method as a baseline (labelled as Bitcoin).

From top to bottom, the evaluation of VaR is presented sequentially in decline, raise and flat phases in the left panels. In the declining phase, all three methods (normal distribution, t-distribution and GA method) underestimate the risk level of the observations, which has a significant negative drift. By contrast, the risk level is overestimated by the three methods in the increasing phase. While in the flat phase, the three methods can only partially cover the baseline curve. None of them can provide a satisfactory performance due to risk evaluation using VaR is unsuitable for skewed distributions, and the log return distribution of three phases are asymmetric,

whose skewness are -0.05, -0.55 and 0.57, respectively.

However, in the right panels, the baseline curve of CVaR can be well fitted by the GA method in the decline phase (top figure) and the t-distribution method in the flat phase (bottom figure), and the GA method can provide a close output to the baseline curve in the raising phase (middle figure).

In general, the GA calibration combined with an ABM and t-distribution fitting method can provide a satisfactory risk prediction for Bitcoin data using the CVaR metric, which is a coherent risk measure and continuous to the confidence level. By contrast, the VaR metric is not a suitable choice, since it may lead to undesirable results for skewed distributions.

Finally, we display the results across a range of statistical metrics that we perform the two-sample tests, such as Kolmogorov-Smirnov test (KS), Anderson-Darling test (AD) and Cramér-von Mises test (CM), to judge whether the return distribution between simulations and Bitcoin price is the same or not. We use the probability distribution fitting approach that the log return of observations is assumed to follow the normal distribution and t distribution, respectively. The maximum likelihood estimators are set as probability parameter inputs to generate random log returns with these two distributions. At the same time, the GA method inputs its optimal solution into the ABM to produce a simulated log return series. In total, there are 9 groups of log returns performed with two sample tests.

In Table 4.2, the information, for instance moments of Bitcoin price, test statistics and the p-values are provided. From the view of three tests (KS, AD and CM), there is no significant different between the Bitcoin price and the simulated price via three methods in all three phases even at 15% confidence level. Different from the situation in calibrating against simulated data in Section 4.4.1, the log return distribution of Bitcoin price has a relatively smaller kurtosis and skewness comparing with simulated observations, under which the two sample tests have less power to distinguish the difference between log return distributions.

4.5 Summary

In this chapter, we applied the GA to calibrate against simulated and Bitcoin data, respectively. The aim is to find a suitable region of parameters, which can be input to the ABM (introduced in Chapter 2) to replicate the simulated observations or Bitcoin price. In order to find the optimal parameters of an ABM to fit the observations, we proposed an objective function, which minimises the linear combinations of moment difference between observed and simulated log returns. The fitness function of the GA can be easily obtained by taking the reciprocal of the objective function, which is used to evaluate the goodness of the GA population. Although it is hard to find a global optimum due to the complexity and stochasticity of our ABM, the GA can provide a region close to optimal value, within which the population have a high quality performance in respect of the objective function.

Furthermore, two popular risk metrics, VaR and CVaR are applied to testify whether we can combine the ABM and GA to predict the future distribution of stock log return given some observations. For comparison, we also produce two more simulated results using normal and t distribution fitting method to calculate their corresponding VaR and CVaR.

For calibrating against simulated data, we found that there is no suitable method to properly evaluate VaRs in the case that observations are generated by a ABM with small values of parameter w. Since the parameter w determines the kurtosis of the simulated observations and a small value indicates a fat tail in log return distribution. The GA method combined with the ABM performs poorly, because it can not sufficiently capture the feature of observations, whose parameters only partially calibrate the observation in volatility. By contrast, t-distribution method has the best performance in evaluating CVaR, while the GA method perform badly due to a poor tail modelling. Moreover, three two-sample tests (KS, AD and CM test) are performed to judge whether the return distribution between simulations and observations are the same or not. There is no significant difference between observed log returns and simulated log returns via t-distribution and GA method even at 10% confidence level.

For calibrating against Bitcoin price, we introduced a parameter b which controls initial money given to each agent in the ABM. Although we can not properly evaluate

the VaRs for the Bitcoin price because it is unstable for skewed distributions, the GA and t-distribution method can well fit the CVaR curve if we set the CVaRs calculated by historical Bitcoin as a baseline. It is noteworthy that there is no significant different between the log-return of Bitcoin price and the simulated price in all cases we summarised in Table 4.2 by performing KS, AD and CM test.

Cases	Skewness	Kurtosis	Groups	Two-Sample Tests					
				KS	AD	CM	KS	AD	CM
				Test Statistic			p-value		
			Obs and Normal-dist	0.245	1.623	0.410	0.11	0.07	0.07
$w = 0$	1.738	10.017	Obs and t-dist	0.122	-0.236	0.082	0.86	> 0.25	0.70
			Obs and GA	0.143	-0.758	0.079	0.71	> 0.25	0.71
			Obs and Normal-dist	0.245	1.408	0.345	0.11	0.09	0.10
$w = 0.1$	-0.572	6.996	Obs and t-dist	0.122	-0.135	0.107	0.86	> 0.25	0.56
			Obs and GA	0.143	-0.347	0.117	0.71	> 0.25	0.52
			Obs and Normal-dist	0.286	2.114	0.453	0.04	0.04	0.05
$w = 0.3$	1.293	5.264	Obs and t-dist	0.204	0.842	0.246	0.26	0.15	0.20
			Obs and GA	0.184	0.119	0.203	0.38	> 0.25	0.27
			Obs and Normal-dist	0.224	0.838	0.302	0.17	0.15	0.14
$w = 0.5$	-1.257	3.700	Obs and t-dist	0.102	-0.697	0.057	0.96	> 0.25	0.85
			Obs and GA	0.163	-0.559	0.111	0.54	> 0.25	0.54
			Obs and Normal-dist	0.143	-0.376	0.104	0.70	> 0.25	0.58
$w = 0.8$	0.151	1.296	Obs and t-dist	0.122	-0.641	0.059	0.86	> 0.25	0.84
			Obs and GA	0.102	-0.770	0.043	0.96	> 0.25	0.93
			Obs and Normal-dist	0.061	-1.196	0.014	1.00	> 0.25	1.00
$w = 1$	0.236	0.366	Obs and t-dist	0.163	-0.103	0.143	0.54	> 0.25	0.42
			Obs and GA	0.082	-1.088	0.026	1.00	> 0.25	1.00

Table 4.1: Normality tests for two sample paths. Six scenarios are presented that observations are generated with different values of w. The tests, such as Kolmogorov-Smirnov test (KS), Anderson-Darling test (AD) and Cramér-von Mises test (CM) are conducted in three groups of log-returns for each scenario.

Mean	Skewness	Kurtosis	Groups	Two-Sample Tests					
				KS	AD	CM	KS	AD	CM
				Test Statistic				p-value	
			Obs and Normal-dist	0.143	-0.580	0.098	0.71	> 0.25	0.61
-0.010	-0.049	0.497	Obs and t-dist	0.102	-0.911	0.048	0.96	> 0.25	0.90
			Obs and GA	0.130	-0.200	0.133	0.57	> 0.25	0.45
			Obs and Normal-dist	0.195	0.710	0.259	0.25	0.168	0.18
0.008	-0.551	1.147	Obs and t-dist	0.122	-0.740	0.056	0.87	> 0.25	0.85
			Obs and GA	0.130	-0.364	0.092	0.61	> 0.25	0.64
			Obs and Normal-dist	0.163	-0.325	0.121	0.54	> 0.25	0.50
-0.001	0.574	1.069	Obs and t-dist	0.143	-0.269	0.111	0.71	> 0.25	0.54
			Obs and GA	0.134	-0.277	0.143	0.55	> 0.25	0.42

Table 4.2: Two sample Kolmogorov-Smirnov, Anderson-Darling Cramér-von Mises results against three groups of log-returns, which are observations and a fitted normal distribution (group 1), observations and a fitted t-distribution (group 2) and observations and the GA calibration (group 3).

Chapter 5

General Conclusions

This thesis has introduced a framework of using agent-based models with heterogeneous agents to produce an artificial financial market. We explored the herding effect of an endogenous imitation mechanism and verified a set of stylised empirical facts. We also introduced trend-following agents using the Bayesian learning method to track the asset return. Moreover, a procedure of model calibration have been provided. Here, we summarise the main findings and discuss avenues for future research.

5.1 Conclusion

Chapter 2 started by experimenting with a model proposed by Tedeschi *et al.* (2009) [52], in which an order-driven market can be built on some essential elements, such as the trading mechanism, the behaviour of typical agents and a communication network system among agents. By involving an endogenous mechanism of imitation, the herding effect can be generated that wealthy agents are more likely to be imitated by others, causing a larger fluctuation in price. Due to the model being a closed system, the price movement will be restricted to oscillate around a constant P^{mean}, which is the ratio of the total money supply to total share supply in the system. The model verified a set of empirical facts that stock returns exhibit a leptokurtic, asymmetric and fat-tailed distribution and lack auto-correlation.

We have contributed to implementing the parameter sensitivity analysis based on the one-factor-at-a-time method, which is prior to the calibration process in Chapter 4 as a descriptive or in-sample exercise identifying the vital parameter for determining

the statistical properties of the model. Based on the analysis, we found that a low bid-ask spread of simulated price can be obtained with a hybrid structure of agent expectation, where more market liquidity can be created. The herding effect is more likely to be observed when the agent's expectation mainly accounts for imitation behaviour. Within a suitable range, the scale parameter A of the agent's specific volatility shows a linear relationship against the first and second moment of the bid-ask spread. Moreover, additional participants will not significantly influence the bid-ask spread when the market has sufficient agents.

In Chapter 3, we first reviewed a portfolio choice problem proposed by Bismuth *et al.* [53] (2019). The ABM was extended by introducing trend-following agents who aim to maximise their utility function by solving a portfolio choice problem. The Agents are assumed to use the Bayesian learning method to track the risky asset return and place the limit orders to achieve optimal strategy goals. The simulation results were presented under the case that agents have a CRRA utility function with $\gamma = 1$, which makes their optimal strategy θ^* proportional to their learning beliefs. We found that for the fixed belief case, the price movement is similar to the ABM with noisy agents that oscillate around a constant dominating by the average optimal risky asset proportion of agents $\tilde{\theta^*}$. While agents are equipped with learning techniques, a positive feedback loop was observed that the optimal strategy guides agents to reduce their investment in a risky asset when $\theta^* < \theta$. Then a substantial amount of ask orders pull down the price leading to a lower optimal proportion θ^*. An extended parameter sensitivity analysis showed how to choose a suitable initial range of parameters for the model simulation. Any wrong choice of the initial parameter will cause abnormal simulation results, such as a flat price movement due to no completed transaction and a positive feedback loop in price. The flat price movement can be explained by the reasons that: (1) agents beliefs locate in a small range causing the maximum distance of limit order price to be smaller than the minimum tick size of stock price when μ_β is very tiny; (2) agents beliefs are widely distributed that the best bid and best ask order will never crossover with each other to make a transaction, when μ_β is relatively large. While the positive feedback loop is more likely to be viewed with a higher value of v_0, where agents learn faster about the asset return.

Chapter 4 was motivated by the fact that the complexity and stochasticity of an

ABM model do not allow us to find a global optimum. Therefore, we have contributed to using a genetic algorithm to calibrate an ABM, which finds a region close to an optimal value instead of a single value. Moreover, We first attempted to use the calibration results to produce a future empirical distribution of log-return, which is practical to calculate risk metrics for a risk management purpose. We found that the calibrating with a GA method does not outperform the traditional probability fitting approaches in the risk prediction. However, an ABM calibrating against the Bitcoin price with a GA method performed well in producing an empirical return distribution.

5.2 Future Work

There is huge scope for extending the methods in this thesis given the number of parameters and assumptions required to build up a functioning limit order market. Our first thought relates to the parameter sensitivity analysis implemented in Chapter 2 and 3, where our one-factor-at-a-time analysis only roughly evaluates the partial derivative, and more accurate analysis requires a smaller step size in parameter values. Instead, we could change it to Sobol's method proposed by Salteli *et. al.* [81], which is model-free, and the sensitivity indices are defined as the ratio of the partial variance to the total variance. The accuracy of the estimated sensitivity indices can therefore be obtained by using a bootstrap method.

Next, on reviewing Chapter 3, we were unable to adequately test the model where agents have the CARA utility function or the CRRA utility function with $\gamma \neq 1$. More investigations can be implemented to identify the statistical properties of model outputs with these settings. Furthermore, input validation can be a further extension since a slight deviation of initial parameter inputs may lead to chaotic dynamics of simulated outputs. Apart from parameter calibration, an exploration of the parameter space can be helpful to evaluate the influences of different parameters on the dynamics of model outputs. The works related to investigating the robustness of model simulations under different parameter settings can be found in [82] and [83].

Another point raised in Chapter 4 is the ability of ABM techniques to contribute to risk analysis. If we can find an optimal method to evaluate the VaR and CVaR, we can work on an optimisation problem that minimise the VaR and CVaR of the portfolio

return under the constraint that the expected return exceeds some pre-specified level.

Finally, we only calibrated an ABM from Chapter 2 in Chapter 4, one obvious extension would be to calibrate an ABM with the trend-following agents of Chapter 3. If the technique can be shown to be reliable, then by calibrating against cryptocurrency data with a GA method and trend following agents, we could ask questions such as:

- What are the likely risk aversion parameters of agents in the market?
- What proportion of agent are implementing trend following strategies in the market?
- Are there any objective differences between cryptocurrency markets and more established equity markets?
- How likely is it a feedback loop will occur and the price will crash?

The field of ABM models is an exciting area with many avenues left unexplored.

Bibliography

[1] William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

[2] John Lintner. Security prices, risk, and maximal gains from diversification. *The journal of finance*, 20(4):587–615, 1965.

[3] Robert C Merton. An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, pages 867–887, 1973.

[4] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, pages 77–105, 1992.

[5] Franziska Klügl and Ana LC Bazzan. Agent-based modeling and simulation. *Ai Magazine*, 33(3):29–29, 2012.

[6] Mark Paddrik, Roy Hayes, Andrew Todd, Steve Yang, Peter Beling, and William Scherer. An agent based model of the e-mini s&p 500 applied to flash crash analysis. In *Computational Intelligence for Financial Engineering & Economics (CIFEr), 2012 IEEE Conference on*, pages 1–8. IEEE, 2012.

[7] Patrick De Fontnouvelle. Information dynamics in financial markets. *Macroeconomic Dynamics*, 4(2):139–169, 2000.

[8] Bryan R Routledge. Genetic algorithm learning to choose and use information. *Macroeconomic dynamics*, 5(02):303–325, 2001.

[9] Gary S Becker. Irrational behavior and economic theory. *Journal of political economy*, 70(1):1–13, 1962.

[10] Vernon L Smith. An experimental study of competitive market behavior. *Journal of political economy*, 70(2):111–137, 1962.

[11] Dhananjay K Gode and Shyam Sunder. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of political economy*, 101(1):119–137, 1993.

[12] Marco Bartolozzi. A multi agent model for the limit order book dynamics. *The European Physical Journal B*, 78(2):265–273, 2010.

[13] Lars Peter Hansen and Kenneth J Singleton. Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of political economy*, 91(2):249–265, 1983.

[14] Shu-Heng Chen and Chia-Hsuan Yeh. Evolving traders and the business school with genetic programming: A new architecture of the agent-based artificial stock market. *Journal of Economic Dynamics and Control*, 25(3-4):363–393, 2001.

[15] Shu-Heng Chen and Chia-Hsuan Yeh. On the emergent properties of artificial stock markets: the efficient market hypothesis and the rational expectations hypothesis. *Journal of Economic Behavior & Organization*, 49(2):217–239, 2002.

[16] Marco Raberto, Silvano Cincotti, Sergio M Focardi, and Michele Marchesi. Agent-based simulation of a financial market. *Physica A: Statistical Mechanics and its Applications*, 299(1-2):319–327, 2001.

[17] W Brian Arthur, John H Holland, Blake LeBaron, Richard Palmer, and Paul Tayler. Asset pricing under endogenous expectations in an artificial stock market. *The economy as an evolving complex system II*, 27, 1996.

[18] Blake LeBaron, W Brian Arthur, and Richard Palmer. Time series properties of an artificial stock market. *Journal of Economic Dynamics and control*, 23(9-10):1487–1516, 1999.

[19] Richard G Palmer, W Brian Arthur, John H Holland, Blake LeBaron, and Paul Tayler. Artificial economic life: a simple model of a stockmarket. *Physica D: Nonlinear Phenomena*, 75(1-3):264–274, 1994.

[20] Thomas Landes and Otto Loistl. Complexity models in financial markets. *Applied Stochastic Models and Data Analysis*, 8(3):209–228, 1992.

[21] Michael Youssefmir, Bernardo A Huberman, and Tad Hogg. Bubbles and market crashes. *Computational Economics*, 12(2):97–114, 1998.

[22] J Doyne Farmer and Shareen Joshi. The price dynamics of common trading strategies. *Journal of Economic Behavior & Organization*, 49(2):149–171, 2002.

[23] Rui Carvalho. The dynamics of the linear random farmer model. *arXiv preprint cond-mat/0107150*, 2001.

[24] Masanao Aoki. Open models of share markets with two dominant types of participants. *Journal of economic behavior & organization*, 49(2):199–216, 2002.

[25] William J Baumol. Speculation, profitability, and stability. *The Review of Economics and Statistics*, pages 263–271, 1957.

[26] EC Zeeman. Differential equations for the heartbeat and nerve impulse. In *Dynamical systems*, pages 683–741. Elsevier, 1973.

[27] E Christopher Zeeman. On the unstable behaviour of stock exchanges. *Journal of mathematical economics*, 1(1):39–49, 1974.

[28] Richard H Day and Weihong Huang. Bulls, bears and market sheep. *Journal of Economic Behavior & Organization*, 14(3):299–329, 1990.

[29] Jeffrey A Frankel and Kenneth Froot. The dollar as speculative bubble: a tale of fundamentalists and chartists. *NBER Working Paper*, (w1854), 1986.

[30] Jeffrey A Frankel, Kenneth A Froot, et al. Understanding the us dollar in the eighties: the expectations of chartists and fundamentalists. *Economic record*, 62(1):24–38, 1986.

[31] Paul De Grauwe and Hans Dewachter. A chaotic model of the exchange rate: the role of fundamentalists and chartists. *Open economies review*, 4(4):351–379, 1993.

[32] Blake LeBaron. Technical trading rules and regime shifts in foreign exchange. *Advanced trading rules*, pages 5–40, 1998.

[33] Philip Treleaven, Michal Galas, and Vidhi Lalchand. Algorithmic trading review. *Communications of the ACM*, 56(11):76–85, 2013.

[34] Álvaro Cartea, Sebastian Jaimungal, and Damir Kinzebulatov. Algorithmic trading with learning. *International Journal of Theoretical and Applied Finance*, 19(04):1650028, 2016.

[35] Fabien Guilbaud and Huyen Pham. Optimal high-frequency trading with limit and market orders. *Quantitative Finance*, 13(1):79–94, 2013.

[36] Alvaro Cartea and Sebastian Jaimungal. Modelling asset prices for algorithmic and high-frequency trading. *Applied Mathematical Finance*, 20(6):512–547, 2013.

[37] Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci. Trading strategies within the edges of no-arbitrage. *International Journal of Theoretical and Applied Finance*, 21(03):1850025, 2018.

[38] Sandrine Jacob Leal, Mauro Napoletano, Andrea Roventini, and Giorgio Fagiolo. Rock around the clock: An agent-based model of low-and high-frequency trading. *Journal of Evolutionary Economics*, 26(1):49–76, 2016.

[39] Burton G Malkiel and Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.

[40] Paul Windrum, Giorgio Fagiolo, and Alessio Moneta. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation*, 10(2):8, 2007.

[41] Claudia Werker and Thomas Brenner. Empirical calibration of simulation models. Technical report, Papers on Economics and Evolution, 2004.

[42] Giorgio Fagiolo, Mattia Guerini, Francesco Lamperti, Alessio Moneta, and Andrea Roventini. Validation of agent-based models in economics and finance. Technical report, LEM Working Paper Series, 2017.

[43] Christian Gourieroux, Alain Monfort, and Eric Renault. Indirect inference. *Journal of applied econometrics*, 8(S1):S85–S118, 1993.

[44] Jakob Grazzini, Matteo G Richiardi, and Mike Tsionas. Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control*, 77:26–47, 2017.

[45] Robert Ernest Marks. Validating simulation models: a general framework and four applied examples. *Computational economics*, 30(3):265–290, 2007.

[46] Sandrine Jacob Leal and Mauro Napoletano. Market stability vs. market resilience: Regulatory policies experiments in an agent-based model with low-and high-frequency trading. *Journal of Economic Behavior & Organization*, 2017.

[47] Maria Cristina Recchioni, Gabriele Tedeschi, and Mauro Gallegati. A calibration procedure for analyzing stock price dynamics in an agent-based framework. *Journal of Economic Dynamics and Control*, 60:1–25, 2015.

[48] William A Brock and Cars H Hommes. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic dynamics and Control*, 22(8-9):1235–1274, 1998.

[49] Simone Alfarano, Thomas Lux, and Friedrich Wagner. Estimation of a simple agent-based model of financial markets: An application to australian stock and foreign exchange data. *Physica A: Statistical Mechanics and its Applications*, 370(1):38–42, 2006.

[50] Alex Rogers and Peter Von Tessin. Multi-objective calibration for agent-based models. 2004.

[51] Annalisa Fabretti. On the problem of calibrating an agent based model for financial markets. *Journal of Economic Interaction and Coordination*, 8(2):277–293, 2013.

[52] Gabriele Tedeschi, Giulia Iori, and Mauro Gallegati. The role of communication and imitation in limit order markets. *The European Physical Journal B*, 71(4):489, 2009.

[53] Alexis Bismuth, Olivier Guéant, and Jiang Pu. Portfolio choice, portfolio liquidation, and portfolio transition under drift uncertainty. *Mathematics and Financial Economics*, 13(4):661–719, 2019.

[54] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[55] Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. 2001.

[56] John Y Campbell, John J Champbell, John W Campbell, Andrew W Lo, Andrew W Lo, and A Craig MacKinlay. *The econometrics of financial markets*. princeton University press, 1997.

[57] Rama Cont, Marc Potters, and Jean-Philippe Bouchaud. Scaling in stock market data: stable laws and beyond. In *Scale invariance and beyond*, pages 75–85. Springer, 1997.

[58] Rama Cont. *Statistical finance: empirical study and theoretical modeling of price variations in financial markets*. PhD thesis, PhD thesis, Universite de Paris XI, 1998.

[59] Adrian Pagan. The econometrics of financial markets. *Journal of empirical finance*, 3(1):15–102, 1996.

[60] Eugene F Fama. Efficient capital markets: Ii. *The journal of finance*, 46(5):1575–1617, 1991.

[61] Portfolio Selection and E Diversification. Harry markowitz. *The Journal of Finance*, 7(1):77–91, 1952.

[62] Paul A Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *The review of economics and statistics*, pages 239–246, 1969.

[63] Robert C Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, pages 247–257, 1969.

[64] Robert C Merton. Optimum consumption and portfolio rules in a continuous-time model. In *Stochastic Optimization Models in Finance*, pages 621–661. Elsevier, 1975.

[65] Ioannis Karatzas and Xiaoliang Zhao. Bayesian adaptive portfolio optimization. *Option pricing, interest rates and risk management*, pages 632–669, 2001.

[66] Ahmed Fawzy Gad. Pygad: An intuitive genetic algorithm python library. *arXiv preprint arXiv:2106.06158*, 2021.

[67] R Tyrrell Rockafellar, Stanislav Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2(3):21–41, 2000.

[68] John L Hodges. The significance probability of the smirnov two-sample test. *Arkiv för Matematik*, 3(5):469–486, 1958.

[69] Fritz W Scholz and Michael A Stephens. K-sample anderson–darling tests. *Journal of the American Statistical Association*, 82(399):918–924, 1987.

[70] Theodore W Anderson. On the distribution of the two-sample cramer-von mises criterion. *The Annals of Mathematical Statistics*, pages 1148–1159, 1962.

[71] Shaen Corbet, Brian Lucey, Andrew Urquhart, and Larisa Yarovaya. Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, 62:182–199, 2019.

[72] Andrew Urquhart. The inefficiency of bitcoin. *Economics Letters*, 148:80–82, 2016.

[73] David Vidal-Tomás and Ana Ibañez. Semi-strong efficiency of bitcoin. *Finance Research Letters*, 27:259–265, 2018.

[74] Wang Chun Wei. Liquidity and market efficiency in cryptocurrencies. *Economics Letters*, 168:21–24, 2018.

[75] Eng-Tuck Cheah and John Fry. Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin. *Economics letters*, 130:32–36, 2015.

[76] Shaen Corbet, Brian Lucey, and Larisa Yarovaya. Datestamping the bitcoin and ethereum bubbles. *Finance Research Letters*, 26:81–88, 2018.

[77] Shaen Corbet, Andrew Meegan, Charles Larkin, Brian Lucey, and Larisa Yarovaya. Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, 165:28–34, 2018.

[78] Ladislav Kristoufek. Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific reports*, 3(1):1–7, 2013.

[79] Elie Bouri, Syed Jawad Hussain Shahzad, and David Roubaud. Co-explosivity in the cryptocurrency market. *Finance Research Letters*, 29:178–183, 2019.

[80] Antonis Ballis and Konstantinos Drakos. Testing for herding in the cryptocurrency market. *Finance Research Letters*, 33:101210, 2020.

[81] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[82] Isabelle Salle and Murat Yıldızoğlu. Efficient sampling and meta-modeling for computational economic models. *Computational Economics*, 44(4):507–536, 2014.

[83] Leonardo Bargigli, Luca Riccetti, Alberto Russo, and Mauro Gallegati. Network calibration and metamodeling of a financial accelerator agent based model. *Journal of Economic Interaction and Coordination*, 15(2):413–440, 2020.

Appendix A

Proofs of propositions in Chapter 3

In this section, we will provide the proofs of Proposition 1 - 8 introduced in Section 3.3 - 3.5.1.

Proposition 1. *Let $t \in R_+$, given $\mathcal{F}_t^P$, μ has a conditional distribution with mean β_t and variance v_t^2, where*

$$\beta_t = \frac{\sigma^2}{\sigma^2 + v_0^2 t}\beta_0 + \frac{v_0^2}{\sigma^2 + v_0^2 t}(\mu_t + \sigma W_t),$$

and

$$v_t^2 = \frac{\sigma^2 v_0^2}{\sigma^2 + v_0^2 t}.$$

Proof. For $0 < t_1 < \ldots < t_n$, $(\mu, \mu t_1 + \sigma W_{t_1}, \ldots, \mu t_n + \sigma W_{t_n})$ is a Gaussian vector with variance matrix

$$\left(\begin{array}{c|c} v_0^2 & L \\ \hline L' & \Theta \end{array}\right)$$

where

$$L = (v_0^2 t_1, \ldots, v_0^2 t_n)$$

and

$$\Theta_{ij} = v_0^2 t_i t_j + \sigma^2 \inf(t_i, t_j).$$

The distribution of μ given $(S_{t_1}, \ldots, S_{t_n})$ is the distribution of μ given $(\mu t_1 + \sigma W_{t_1}, \ldots, \mu t_n + \sigma W_{t_n})$. It is Gaussian $N(\beta^{t_1,\ldots,t_n}, {v^{t_1,\ldots,t_n}}^2)$ with

$$\beta^{t_1,\ldots,t_n} = \beta_0 + L\Theta^{-1}\begin{pmatrix} (\mu - \beta_0)t_1 + \sigma W_{t_1} \\ \vdots \\ (\mu - \beta_0)t_n + \sigma W_{t_n} \end{pmatrix},$$

and

$$v^{t_1,\ldots,t_n 2} = v_0^2 - L\Theta L^{'}$$

But

$$L\Theta^{-1} = \left(0, \ldots, 0, \frac{v_0^2}{\sigma^2 + v_0^2 t_n}\right).$$

Therefore,

$$\beta^{t_1,\ldots,t_n} = \beta_0 + \frac{v_0^2}{\sigma^2 + v_0^2 t_n}((\mu - \beta_0)t_n + \sigma W_{t_n}) = \frac{\sigma^2}{\sigma^2 + v_0^2 t_n}\beta_0 + \frac{v_0^2}{\sigma^2 + v_0^2 t_n}(\mu t_n + \sigma W_{t_n})$$

and

$$v^{t_1,\ldots,t_n 2} = v_0^2 - \frac{v_0^4 t_n}{\sigma^2 + v_0^2 t_n} = \frac{\sigma^2 v_0^2}{\sigma^2 + v_0^2 t_n}.$$

By a monotone class argument, we have therefore that, for $t \geq 0$, the distribution of μ given $\mathcal{F}_t^S$ is Gaussian with mean

$$\frac{\sigma^2}{\sigma^2 + v_0^2 t}\beta_0 + \frac{v_0^2}{\sigma^2 + v_0^2 t}(\mu t + \sigma W_t) =: \beta_t$$

and variance

$$\frac{\sigma^2 v_0^2}{\sigma^2 + v_0^2 t} =: v_t^2.$$

□

Proposition 2. $\left(\hat{W}_t\right)_{t \in R_+}$ *is a Wiener process adapted to* $(\mathcal{F}_t^S)_{t \in R_+}$.

Proof. For proving this result, we use the Levy's characterisation of a Brownian motion. Let $t \in R_+$. By definition, we have

$$\hat{W}_t = \frac{1}{\sigma}\left(\log\left(\frac{S_t}{S_0}\right) + \frac{1}{2}\sigma^2 t\right) - \int_0^t \frac{\beta_s}{\sigma} ds,$$

hence the $\mathcal{F}_t^S$-measurability of $\hat{W}_t$. Let $s, t \in R_+$, with $s < t$

$$E\left[\hat{W}_t - \hat{W}_s | \mathcal{F}_s^S\right] = E\left[W_t - W_S | \mathcal{F}_s^S\right] + E\left[\int_s^t \frac{\mu - \beta_{t'}}{\sigma} dt' | \mathcal{F}_s^S\right].$$

For the first term, the increment $W_t - W_s$ is independent of $\mathcal{F}_s^W$ and independent of μ. Therefore, it is independent of $\mathcal{F}_s^S$ and we have

$$\begin{aligned} E\left[\int_s^t \frac{\mu - \beta_{t'}}{\sigma} dt' | \mathcal{F}_s^S\right] &= \int_s^t E\left[\frac{\mu - \beta_{t'}}{\sigma} | \mathcal{F}_s^S\right] dt' \\ &= \int_s^t E\left[E\left[\frac{\mu - \beta_{t'}}{\sigma} | \mathcal{F}_{t'}^S\right] | \mathcal{F}_s^S\right] dt' \\ &= 0, \end{aligned}$$

by definition of $\beta_{t'}$.

We obtain that $\left(\hat{W}_t\right)_t$ is an $\mathcal{F}^S$-martingale. Since $\hat{W}$ has continuous paths and $d\left\langle \hat{W}_t \right\rangle = dt$, we conclude that $\hat{W}$ is an $\mathcal{F}^S$-Brownian motion. □

Proposition 3. *Suppose there exists $\varphi \in \mathcal{C}^{1,2}\left([0,T]\times R\right)$ satisfying*

$$-\partial_t\varphi(t,\beta) - \frac{1}{2}\sigma^2 g^2(t)\partial^2_{\beta\beta}\varphi(t,\beta) - \frac{(\beta-r)^2}{2\gamma\sigma^2} + g(t)(\beta-r)\partial_\beta\varphi(t,\beta) = 0, \tag{A.1}$$

with the terminal condition

$$\varphi(T,\beta) = 0. \tag{A.2}$$

Then the u defined by

$$u(t,V,\beta) = -\exp\left[-\gamma\left(e^{r(T-t)}V + \varphi(t,\beta)\right)\right]. \tag{A.3}$$

is solution of the HJB equation (A.1) with terminal condition (A.2), and the supremum of the HJB equation

$$\begin{aligned} -\partial_t u - \frac{1}{2}\sigma^2 g(t)^2\partial^2_{\beta\beta}u - \sup_{M\in\mathcal{A}_t}\{((\beta-r)M_t + rV)\partial_V u + \\ \frac{1}{2}\sigma^2 M^2\partial^2_{VV}u + \sigma^2 M g(t)\partial^2_{\beta V}u\} = 0, \end{aligned} \tag{A.4}$$

can be obtained at

$$M^* = e^{-r(T-t)}\left(\frac{\beta-r}{\gamma\sigma^2} - g(t)\partial_\beta\varphi(t,\beta)\right). \tag{A.5}$$

Proof. Let us consider ϕ solution (A.1) of with terminal condition (A.2). For u defined by (A.3), we have

$$\begin{aligned} &-\partial_t u - \frac{1}{2}\sigma^2 g^2\partial^2_{\beta\beta}u - \sup_M\{((\beta-r)M_t + rV)\partial_V u + \frac{1}{2}\sigma^2 M^2\partial^2_{VV}u + \sigma^2 M g\partial^2_{\beta V}u\} \\ =&-\gamma u\Big(-\partial_t\varphi + re^{r(T-t)}V + \frac{1}{2}\sigma^2 g^2\left(\gamma\left(\partial_\beta\varphi\right)^2 - \partial^2_{\beta\beta}\varphi\right) - \sup_M\Big\{((\beta-r)M + rV)e^{r(T-t)} \\ &-\frac{\gamma}{2}\sigma^2\left(e^{r(T-t)}M\right)^2 - \gamma\sigma^2 g\left(e^{r(T-t)}M\right)\partial_\beta\varphi\Big\}\Big) \\ =&-\gamma u\Big(-\partial_t\varphi + \frac{1}{2}\sigma^2 g^2\left(\gamma\left(\partial_\beta\varphi\right)^2 - \partial^2_{\beta\beta}\varphi\right) - \sup_M\Big\{(\beta-r)e^{r(T-t)}M \\ &-\frac{\gamma}{2}\sigma^2\left(e^{r(T-t)}M\right)^2 - \gamma\sigma^2 g\left(e^{r(T-t)}M\right)\partial_\beta\varphi\Big\}\Big). \end{aligned}$$

Now,

$$\begin{aligned} &\sup_M\Big\{(\beta-r)e^{r(T-t)}M - \frac{\gamma}{2}\sigma^2\left(e^{r(T-t)}M\right)^2 - \gamma\sigma^2 g\left(e^{r(T-t)}M\right)\partial_\beta\varphi\Big\} \\ =&\sup_M\Big\{(\beta-r)\tilde{M} - \frac{\gamma}{2}\sigma^2\tilde{M} - \gamma\sigma^2 g\tilde{M}\partial_\beta\varphi\Big\}. \end{aligned}$$

The supremum is reached at

$$\begin{aligned}\tilde{M}^* = M^* e^{r(T-t)} &= \frac{(\beta - r) - \gamma\sigma^2 g \partial_\beta \varphi}{\gamma\sigma^2} \\ &= \frac{(\beta - r)}{\gamma\sigma^2} - g\partial_\beta\varphi,\end{aligned}$$

and

$$\begin{aligned}&\sup_M \Big\{ (\beta - r) e^{r(T-t)} M - \frac{\gamma}{2}\sigma^2 \left(e^{r(T-t)} M\right)^2 - \gamma\sigma^2 g \left(e^{r(T-t)} M\right) \partial_\beta \varphi \Big\} \\ =&\frac{\left((\beta - r) - \gamma\sigma^2 g \partial_\beta \varphi\right)^2}{2\gamma\sigma^2}.\end{aligned}$$

By using this expression, we obtain

$$\begin{aligned}&- \partial_t u - \frac{1}{2}\sigma^2 g^2 \partial^2_{\beta\beta} u - \sup_M \{((\beta - r) M_t + rV)\partial_V u + \frac{1}{2}\sigma^2 M^2 \partial^2_{VV} u + \sigma^2 M g \partial^2_{\beta V} u\} \\ =&-\gamma u \Big(- \partial_t \varphi + \frac{1}{2}\sigma^2 g^2 \left(\gamma \left(\partial_\beta \varphi\right)^2 - \partial^2_{\beta\beta}\varphi\right) - \frac{\left((\beta - r) - \gamma\sigma^2 g \partial_\beta \varphi\right)^2}{2\gamma\sigma^2} \Big) \\ =&-\gamma u \Big(- \partial_t \varphi - \frac{1}{2}\sigma^2 g^2 \partial^2_{\beta\beta}\varphi - \frac{(\beta - r)^2}{2\gamma\sigma^2} + (\beta - r) g \partial_\beta \varphi \Big) \\ =&0,\end{aligned}$$

by definition of φ.

As far as the terminal condition if concerned, it is straightforward to verify that u satisfies the terminal condition

$$u(t, V, \beta) = -\exp(-\gamma V). \tag{A.6}$$

□

Proposition 4. *Suppose that $(a, b) \in (\mathcal{C}^1([0, T]))^2$ satisfies the following system of ODEs:*

$$\begin{cases} a'(t) + \frac{1}{2}\sigma^2 g^2(t) b(t) & = 0 \\ b'(t) + \frac{1}{\gamma\sigma^2} - 2g(t)b(t) & = 0, \end{cases} \tag{A.7}$$

with terminal conditions

$$\begin{cases} a(T) & = 0 \\ b(T) & = 0. \end{cases} \tag{A.8}$$

Then the φ defined by

$$\varphi(t, \beta) = a(t) + \frac{1}{2} b(t)(\beta - r)^2, \tag{A.9}$$

satisfies (A.1) with terminal condition (A.2).

Proof. Let us consider a couple $(a, b) \in (\mathcal{C}^1([0,T]))^2$ solution of (A.7) with terminal condition (A.8). If φ is defined by (A.9), then

$$\begin{aligned}
&-\partial_t\varphi(t,\beta) - \frac{1}{2}\sigma^2 g^2(t)\partial^2_{\beta\beta}\varphi(t,\beta) - \frac{(\beta-r)^2}{2\gamma\sigma^2} + (\beta-r)g(t)\partial_\beta\varphi(t,\beta)\\
=&-a^{'}(t) - \frac{1}{2}b^{'}(t)(\beta-r)^2 - \frac{1}{2}\sigma^2 g^2(t)b(t) - \frac{(\beta-r)^2}{2\gamma\sigma^2} + g(t)b(t)(\beta-r)^2\\
=&0,
\end{aligned}$$

by definition of a and b.

As far as the terminal condition is concerned, it is straightforward to verify that φ satisfies the terminal condition (A.2). □

Proposition 5. *For $\gamma \neq 1$, assume there exists $t^{'} \in [0,T)$ and $\varphi \in \mathcal{C}^{1,2}([t^{'},T]\times R)$ satisfying*

$$\begin{aligned}
-\frac{1}{1-\gamma}\partial_t\varphi(t,\beta) - \frac{1}{2(1-\gamma)}\sigma^2 g^2(t)\partial^2_{\beta\beta}\varphi(t,\beta) - \frac{1}{2\gamma(1-\gamma)}\sigma^2 g^2(t)(\partial_\beta\varphi(t,\beta))^2\\
-\frac{1}{\gamma}\frac{(\beta-r)^2}{2\sigma^2} - \frac{1}{\gamma}g(t)(\beta-r)\partial_\beta\varphi(t,\beta) = 0,
\end{aligned} \tag{A.10}$$

with terminal condition

$$\varphi(T,\beta) = 0. \tag{A.11}$$

Then function u defined by

$$u(t,V,\beta) = \frac{\left(e^{r(T-t)}V\right)^{1-\gamma}}{1-\gamma}\exp\left[\varphi(t,\beta)\right] \tag{A.12}$$

is the solution of HJB equation

$$\begin{aligned}
-\partial_t u(t,V,\beta) - \frac{1}{2}\sigma^2 g^2(t)\partial^2_{\beta\beta}u(t,V,\beta) - \sup_\theta\Big\{&((\beta-r)\theta + r)V\partial_V u(t,V,\beta)\\
&+\frac{1}{2}\sigma^2\theta^2V^2\partial^2_{VV}u(t,V,\beta) + \sigma^2 g(t)\theta V\partial^2_{V\beta}u(t,V,\beta)\Big\} = 0
\end{aligned} \tag{A.13}$$

with the terminal condition

$$u(T,V,\beta) = U^\gamma(V) \tag{A.14}$$

on $[t^{'},T]\times R^*_+\times R$. Furthermore, the supremum in (A.13) can be obtained at

$$\theta^* = \frac{\beta-r}{\gamma\sigma^2} + \frac{1}{\gamma}g(t)\partial_\beta\varphi(t,\beta). \tag{A.15}$$

Proof. Let us consider φ solution of (A.10) with terminal condition (A.11). For u defined by (A.12), we have

$$
\begin{aligned}
&- \partial_t u - \frac{1}{2}\sigma^2 g^2 \partial^2_{\beta\beta} u - \sup_\theta \Big\{ ((\beta - r)\theta + r) V \partial_V u + \frac{1}{2}\sigma^2 \theta^2 V^2 \partial^2_{VV} u + \sigma^2 g \theta V \partial^2_{V\beta} u \Big\} \\
=& - u(\partial_t \varphi - (1-\gamma) r) - \frac{1}{2}\sigma^2 g^2 u \left(\partial^2_{\beta\beta}\varphi + (\partial_\beta \varphi) \right) - \sup_\theta \Big\{ \left((\beta - r)\theta + r \right)(1-\gamma) u \\
&- \frac{1}{2}\gamma(1-\gamma)\sigma^2\theta^2 u + \theta \sigma^2 g (1-\gamma) u \partial_\beta \varphi \Big\} \\
=& u(1-\gamma) \Bigg(- \frac{1}{1-\gamma}\partial_t \varphi - \frac{1}{2(1-\gamma)} \sigma g^2 \left(\partial^2_{\beta\beta}\varphi + (\partial_\beta \varphi)^2 \right) \\
&- \sup_\theta \Bigg\{ \left((\beta - r)\theta - \frac{1}{2}\gamma\sigma^2\theta^2 + \theta\sigma^2 g \partial_\beta \varphi \right) \Bigg\} \Bigg).
\end{aligned}
$$

The supremum is reached at

$$\theta^* = \frac{\beta - r}{\gamma \sigma^2} + \frac{1}{\gamma} g \partial_\beta \varphi.$$

By using this expression, we obtain

$$
\begin{aligned}
&- \partial_t u - \frac{1}{2}\sigma^2 g^2 \partial^2_{\beta\beta} u - \sup_\theta \Big\{ ((\beta - r)\theta + r) V \partial_V u + \frac{1}{2}\sigma^2 \theta^2 V^2 \partial^2_{VV} u + \sigma^2 g \theta V \partial^2_{V\beta} u \Big\} \\
=& u(1-\gamma) \left(- \frac{1}{1-\gamma}\partial_t \varphi - \frac{1}{2(1-\gamma)} \sigma^2 g^2 \left(\partial^2_{\beta\beta}\varphi + (\partial_\beta \varphi)^2 \right) - \frac{1}{2\gamma\sigma^2} \left((\beta - r) + \sigma^2 g \partial_\beta \varphi \right)^2 \right) \\
=& u(1-\gamma) \Bigg(- \frac{1}{1-\gamma}\partial_t \varphi - \frac{1}{2(1-\gamma)} \sigma^2 g^2 \left(\partial^2_{\beta\beta}\varphi + (\partial_\beta \varphi)^2 \right) \\
&- \frac{1}{\gamma} \left(\frac{1}{2\sigma^2}(\beta - r)^2 + (\beta - r) g \partial_\beta \varphi + \frac{1}{2}\sigma^2 g^2 (\partial_\beta \varphi)^2 \right) \Bigg) \\
=& u(1-\gamma) \Bigg(- \frac{1}{1-\gamma}\partial_t \varphi - \frac{1}{2(1-\gamma)} \sigma^2 g^2 \partial^2_{\beta\beta}\varphi - \frac{1}{2\gamma(1-\gamma)} \sigma^2 g^2 (\partial_\beta \varphi)^2 \\
&- \frac{1}{\gamma}\frac{(\beta - r)^2}{2\sigma^2} - \frac{1}{\gamma}(\beta - r) g \partial_\beta \varphi \Bigg) \\
=& 0,
\end{aligned}
$$

by definition of φ. Finally, it is evident that u satisfies the terminal condition (A.14).

□

Proposition 6. *For $\gamma \neq 1$, suppose there exists $t' \in [0, T]$ and $(a, b) \in \left(\mathcal{C}^1([t', T]) \right)^2$ satisfying the following system of ODEs*

$$
\begin{cases}
a'(t) + \frac{1}{2}\sigma^2 g^2(t) b(t) & = 0 \\
b'(t) + \frac{1}{\gamma}\sigma^2 g^2(t) b^2(t) + \frac{1-\gamma}{\gamma}\frac{1}{\sigma^2} + 2\frac{1-\gamma}{\gamma} g(t) b(t) & = 0,
\end{cases}
\tag{A.16}
$$

with terminal conditions

$$\begin{cases} a(T) &= 0 \\ b(T) &= 0. \end{cases} \tag{A.17}$$

Then the function φ defined by

$$\varphi(t,\beta) = a(t) + \frac{1}{2}b(t)(\beta - r)^2, \tag{A.18}$$

satisfies (A.10) with terminal condition (A.11).

Proof. Let us consider a couple $(a,b) \in \left(\mathcal{C}^1([t',T])\right)^2$ solution of (6) with terminal condition (A.17). If φ is defined by (A.18), then

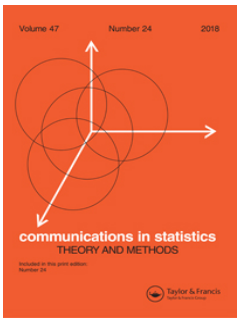
$$\begin{aligned} &- \frac{1}{1-\gamma}\partial_t\varphi(t,\beta) - \frac{1}{2(1-\gamma)}\sigma^2 g^2(t)\partial^2_{\beta\beta}\varphi(t,\beta) \\ &- \frac{1}{2\gamma(1-\gamma)}\sigma^2 g^2(t)(\partial_\beta\varphi(t,\beta))^2 - \frac{1}{\gamma}\frac{(\beta-r)^2}{2\sigma^2} - \frac{1}{\gamma}g(t)(\beta-r)\partial_\beta\varphi(t,\beta) \\ =& - \frac{1}{1-\gamma}a'(t) - \frac{1}{2(1-\gamma)}b'(t)(\beta-r)^2 - \frac{1}{2(1-\gamma)}\sigma^2 g^2(t)b(t) \\ &- \frac{1}{2\gamma(1-\gamma)}\sigma^2 g^2(t)b^2(t)(\beta-r)^2 - \frac{1}{\gamma}\frac{(\beta-r)^2}{2\sigma^2} - \frac{1}{\gamma}g(t)b(t)(\beta-r)^2 \\ =&0, \end{aligned}$$

by definition of a and b.

As far as the terminal condition is concerned, it is straightforward to check that φ satisfies the terminal condition (A.11). □

Appendix B

First year work



Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Simultaneous Bayesian modeling of longitudinal and survival data in breast cancer patients

Ali Azarbar, Yu Wang & Saralees Nadarajah

To cite this article: Ali Azarbar, Yu Wang & Saralees Nadarajah (2019): Simultaneous Bayesian modeling of longitudinal and survival data in breast cancer patients, Communications in Statistics - Theory and Methods, DOI: 10.1080/03610926.2019.1635701

To link to this article: https://doi.org/10.1080/03610926.2019.1635701

Published online: 02 Jul 2019.

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2019.1635701

Simultaneous Bayesian modeling of longitudinal and survival data in breast cancer patients

Ali Azarbar[a], Yu Wang[b], and Saralees Nadarajah[b]

[a]Department of Statistics, Amirkabir University of Technology, Tehran, Iran; [b]School of Mathematics, University of Manchester, Manchester, UK

ABSTRACT

Using simultaneous Bayesian modeling, an attempt is made to analyze data on the size of lymphedema occurring in the arms of breast cancer patients after breast cancer surgery (as the longitudinal data) and the time interval for disease progression (as the time-to-event occurrence). A model based on a multivariate skew t distribution is shown to provide the best fit. This outcome was confirmed by simulation studies too.

ARTICLE HISTORY
Received 11 April 2018
Accepted 19 June 2019

KEYWORDS
Longitudinal and survival data; MCMC method; Simultaneous Bayesian modeling; Time-to-event

2000 MATHEMATICS SUBJECT CLASSIFICATION
49K05; 49K15; 49S05

1. Introduction

The simultaneous modeling of longitudinal and survival data plays a crucial role in the analysis of medical, social, and economic data. For example, in a clinical study aiming at analyzing the data on breast cancer, a biomarker such as the size of a lymph node in the arm of the patient is conceived of as the survival predictor. The putative biomarker is collected for each individual over time until the occurrence of certain events such as patients' refusal, death, or disease recurrence. In most of previously-conducted studies on simultaneous modeling, symmetric distributions, especially the normal distribution, are deployed to model longitudinal responses; however, if the data are not symmetric, this can lead to invalid inferences. In recent years, simultaneous modeling of longitudinal data and time-to-event occurrence has received unprecedented attention. At first, a two-step approach was proposed by Tsiatis et al. (1995). Ye et al. (2008a, 2008b) put forward a two-step regression model to estimate the relationship between longitudinal data and time-to-event occurrence data. Albert and Shih (2010) demonstrated that the bias of the two-step model might be inherent and accordingly, showed how the bias can be decreased using another model. Wulfsohn and Tsiatis (1997) proved that although two-step models have the advantage of computational simplicity, they are plagued with some fundamental problems. Faucett and Thomas (1996) showed that two-step models

CONTACT Saralees Nadarajah mbbsssn2@manchester.ac.uk School of Mathematics, University of Manchester, Manchester, UK.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsta.

use survival data in the modeling of longitudinal processes and, accordingly, might lead to biased results and decrease efficiency of the estimators. Furthermore, the estimated longitudinal measures in the first step are used as constants in the next stage. Hence, this method is characterized by not transferring the uncertainty of the first stage to the second stage. Another method developed to obtain estimates and to make statistical inferences is based on the simultaneous likelihood function for longitudinal and survival data. Deploying this method leads to more exact estimates. Moreover, in this method, the relationship between longitudinal data and time-to-event occurrence is accentuated better. In this method, the parameters of the model are estimated simultaneously and accordingly, in the modeling, uncertainty is taken into account. Simultaneous modeling for the longitudinal and survival data was improved after 1990. Henderson et al. (2000) and Hashemi et al. (2003) conducted a simultaneous modeling using the hidden variable method. Ibrahim et al. (2004), Tsiatis and Davidian (2004), and Wu and Cook (2012), using their own proposed model, attempted to generalize this model. This modeling encompasses innovations in the structure of both longitudinal and survival models. These discussions, especially for longitudinal and survival submodels, involving Bayesian methods, were investigated by Chi and Ibrahim (2007), Bakar et al. (2009), Rizopoulos et al. (2008a, 2008b), Rizopoulos and Ghosh (2011), Rizopoulos (2010, 2013), Baghfalaki et al. (2014) and Baghfalaki and Ganjali (2015). Non-parametric issues to avoid the assumption that random effects are normal were formalized using the EM algorithm method in Song et al. (2002), and Tsiatis and Dividian (2001). Baghfalaki et al. (2014) and Baghfalaki and Ganjali (2015), using heavy-tailed distributions and avoiding commonly-held assumptions such as normality, fitted longitudinal and survival data with common random effects. Issues concerning robust modeling, using the t distribution, were scrutinized in Brown and Ibrahim (2003), Li et al. (2009), and Huang, Chen, and Dagne (2011). Lange et al. (1989) demonstrated that the inaccuracy of the assumption of normality for random effects exerts an adverse effect on the variance of estimates. Given this problem, fitting random effects model in the presence of skewed distributions has attracted researchers' attention. The skew normal family of distributions constitutes an important class of asymmetric distributions. Azzalini (1985) introduced the first version of this family of distributions. Generalizations of this distribution to the multivariate case were provided in Azzalini and Dalla-Valle (1996), Azzalini and Capitanio (1999), Branco and Dey (2001), and Sahu et al. (2003). The applicability of the multivariate normal distribution in mixed-effects models was studied by Arellano-Valle et al. (2007), and Lin and Lee (2008). Furthermore, applicability of this model in the presence of missing data was examined by Lin et al. (2009), Baghfalaki et al. (2014) and Baghfalaki and Ganjali (2015). Choudhary et al. (2014) introduced a general skew-t mixed model that allows different degrees of freedom for random effects. Lu et al. (2017) introduced partially linear mixed effects joint models for skewed and missing longitudinal competing risks outcomes.

In this paper, we analyze longitudinal and survival data with a family of skewed-elliptical multivariate distributions using Bayesian approach. The family is more efficient than the multivariate normal distribution.

The contents of this paper are organized as follows. In Section 2, we introduce skewed-elliptical multivariate distributions. Section 3 introduces

longitudinal and survival models and their notations. Section 4 introduces Bayesian construction for simultaneous modeling of longitudinal and survival data. Section 5 performs a simulation study for checking the proposed model. Section 6 fits the proposed model to a data from breast cancer patients. Discussion and conclusions are given in Section 7.

2. The family of skewed-elliptical multivariate distributions

The elliptical distribution family refers to a set of distributions whose density is an ellipsoidal equation. Suppose $\boldsymbol{\mu} \in R^k$ is the location parameter and $\boldsymbol{\Sigma}$ is a positive-definite matrix with the dimension k and $\mathbf{X}$ is also a k dimensional random vector with the following probability density function:

$$f\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma};g^{(k)}\right) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g^{(k)}\left\{(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}, \mathbf{x} \in R^k \tag{1}$$

where $g^{(k)}$ is the probability density generator function of the random variable of $\mathbf{X}$ defined as

$$g^{(k)}(\cdot) = \frac{\Gamma\left(\frac{k}{2}\right)}{\pi^{\frac{k}{2}}} g(\cdot;k)\left(\int_0^\infty u^{\frac{k}{2}-1} g(u;k)du\right)^{-1}$$

In the case $g(x;k) = \exp(-x/2)$,

$$\begin{aligned} g^{(k)}(x) &= \frac{\Gamma\left(\frac{k}{2}\right)}{\pi^{\frac{k}{2}}} \exp\left(-\frac{x}{2}\right)\left(\int_0^\infty u^{\frac{k}{2}-1}\exp\left(-\frac{u}{2}\right)du\right)^{-1} \\ &= \frac{\Gamma\left(\frac{k}{2}\right)}{\pi^{\frac{k}{2}}} \exp\left(-\frac{x}{2}\right)\left(2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)\right)^{-1} \\ &= \frac{1}{(2\pi)^{\frac{k}{2}}} \exp\left(-\frac{x}{2}\right) \end{aligned}$$

so (1) reduces to

$$f\left(\mathbf{x}\ |\boldsymbol{\mu},\boldsymbol{\Sigma};g^{(k)}\right) = \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{k}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

Other terms in $g^{(k)}$ generate the density in f based on the structure of the combined function.

The skew-elliptical distribution generalizes the elliptical distributions to a family with double flexibility through a skewness parameter. Assume that $\boldsymbol{\epsilon}$ and $\mathbf{Z}$ are k dimensional random vectors and $\boldsymbol{\mu}$ is a m dimensional vector and $\boldsymbol{\Sigma}$ is a positive-definite matrix $k \times k$. In the following, we consider a random vector $\mathbf{X}$ from the elliptical distribution family

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{\epsilon} \\ \mathbf{Z} \end{pmatrix} \sim EL\left(\boldsymbol{\theta}, \boldsymbol{\Omega}; g^{(2k)}\right)$$

where $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu} \\ 0 \end{pmatrix}$, $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma} & 0 \\ 0 & \mathbf{I} \end{pmatrix}$, 0 is the zero matrix and $\mathbf{I}$ is an identity matrix. Set $\mathbf{Y} = \mathbf{DZ} + \boldsymbol{\epsilon}$, where $\mathbf{D} = \text{diag}(d_1, d_2, ..., d_k)$. $\mathbf{Y}|\mathbf{Z} > 0$ extends the class of elliptical

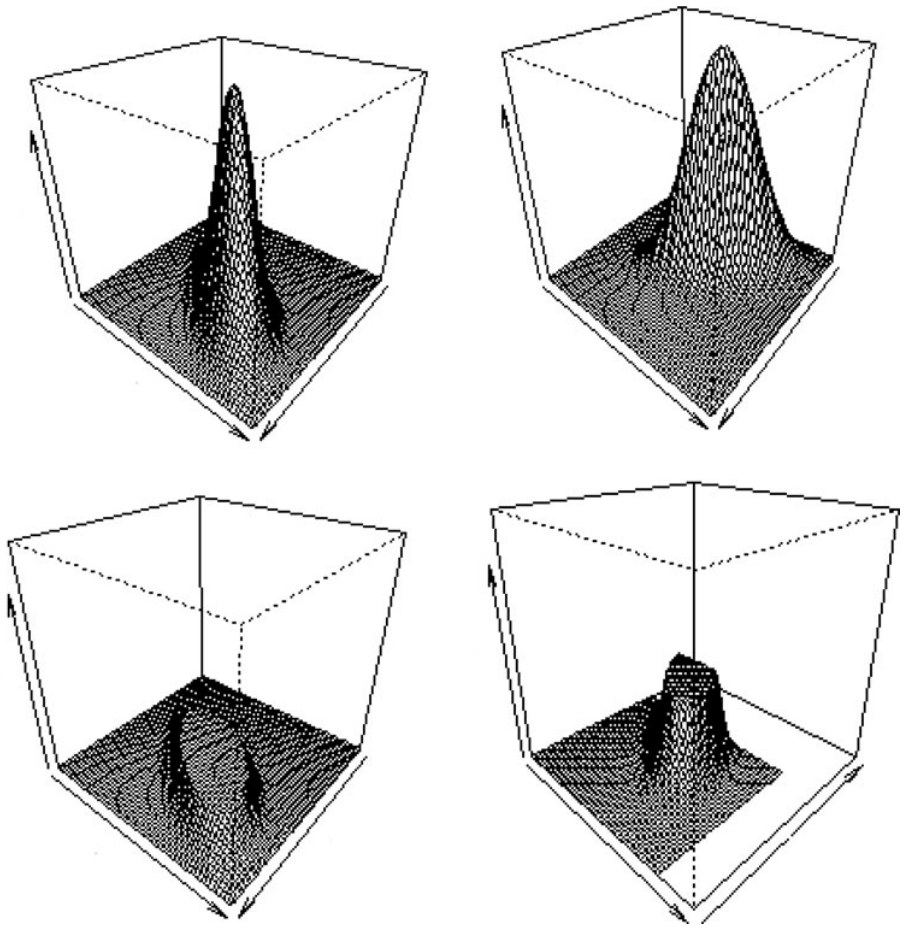


Figure 1. Densities of the bivariate skew normal distribution for $\rho = 0.8, d_1 = 0.7$ and $d_2 = 0.9$ (top left); $\rho = -0.8, d_1 = -0.7$ and $d_2 = -0.9$ (top right); $\rho = 0.4, d_1 = 0.7$ and $d_2 = 0.9$ (bottom left); $\rho = -0.4, d_1 = -0.7$ and $d_2 = -0.9$ (bottom right).

distributions to the skew-elliptical class. If all the elements of the diagonal matrix d_i are positive, the distribution $\mathbf{Y}|\mathbf{Z} > 0$ is right-skewed. However, if d_i are negative, the distribution is left-skewed. Drawing on the definition of conditional density and Bayesian rule, it can be demonstrated that

$$f(\mathbf{y}|\mathbf{Z} > 0) = 2^k f\left(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma} + \mathbf{D}^2; g^{(k)}\right) p(\mathbf{Z} > 0 \mid \mathbf{y}) \tag{2}$$

where $\mathbf{Z} \sim EL(0, \mathbf{I}; g^{(k)})$ and $p(\mathbf{Z} > 0) = 2^{-k}$. One writes $\mathbf{Y} \sim SE(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}; g^{(k)})$. The skew normal distribution is the particular case of (2) for $g(u; m) = \exp(-u/2)$. The skew-t distribution is the particular case of (2) for $g(u; 2m, \nu) = (1 + u/\nu)^{-m-\nu/2}$. For more details, refer to Branco and Dey (2001), and Sahu et al. (2003).

Figure 1 shows density plots of the bivariate skew normal distribution.

3. Model and notation

The initial modeling of the response variables is that besides the error components in the longitudinal sub-model, the random effects in both the longitudinal and survival sub-models have probability distributions. We first assume that $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{in_i})$ is a vector of length n_i to evaluate the response variable of the ith individual. The longitudinal model based on the linear model with mixed effects is

$$\mathbf{Y}_i = \mathbf{X}'_{1i}\boldsymbol{\beta}_1 + \mathbf{Z}'_{1i}\mathbf{b}_{1i} + \boldsymbol{\epsilon}_i$$

for $i = 1, 2, ..., n$. The Cox proportional hazards model with survival time of t_i and censor marker δ_{it} for the ith individual is

$$h(t_i \ |\mathbf{X}_{2i}, \mathbf{Z}_{2i}, \mathbf{b}_{2i}) = h_0(t_i)\exp\left\{\mathbf{X}'_{2i}\boldsymbol{\beta}_2 + \mathbf{Z}'_{2i}\mathbf{b}_{2i}\right\}$$

Replacing non-normal proportional distributions with the deviate nature of the distribution of random effects paves the way for improving the results of estimating models. Accordingly, in the simultaneous modeling based on deploying skew-elliptical distributions for the random-effects model, in the longitudinal and survival submodels, we assume that

$$\begin{aligned}&\mathbf{Y}_i|\boldsymbol{\beta}_1, \boldsymbol{b}_{1i}, \sigma^2_\epsilon \sim N_{n_i}\left(\mathbf{X}'_{1i}\boldsymbol{\beta}_1 + \mathbf{Z}'_{1i}\mathbf{b}_{1i}, \sigma^2_\epsilon\mathbf{I}_{n_i}\right),\\ &\mathbf{b}_{1i}|\boldsymbol{\Psi}_1, \mathbf{D}_{\mathbf{b}_1}; g^{(n_i)}_{\theta_{\mathbf{b}_1}} \sim SE_{q_1}\left(0, \boldsymbol{\Psi}_1, \mathbf{D}_{\mathbf{b}_1}; g^{(n_i)}_{\theta_{\mathbf{b}_1}}\right)\end{aligned} \tag{3}$$

and

$$\mathbf{b}_{2i}|\boldsymbol{\Psi}_2, \mathbf{D}_{\mathbf{b}_2}; g^{(n_i)}_{\theta_{\mathbf{b}_2}} \sim SE_{q_2}\left(0, \boldsymbol{\Psi}_2, \mathbf{D}_{\mathbf{b}_2}; g^{(n_i)}_{\theta_{\mathbf{b}_2}}\right)$$

where $SE_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g_{\boldsymbol{\theta}})$ denotes the q-variable skew elliptical distribution with location parameter of $\boldsymbol{\mu}$, scale matrix of $\boldsymbol{\Sigma}$, skew matrix $\mathbf{D}$ and probability density generator function $g_{\boldsymbol{\theta}}$.

Taking the distributional assumptions introduced in (3), the likelihood function of observations for the simultaneous model is

$$\begin{aligned}L(\boldsymbol{\theta}|\mathbf{y}, \mathbf{b}_1, \mathbf{b}_2, t) &= f\left(\mathbf{y}, \mathbf{b}_1|\boldsymbol{\beta}_1, \sigma^2_\epsilon, \mathbf{D}_{\mathbf{b}_1}, \boldsymbol{\Psi}_1, g_{\theta_{\mathbf{b}_1}}\right) f\left(t, \mathbf{b}_2, \delta_{it}|\boldsymbol{\beta}_2, \mathbf{D}_{\mathbf{b}_1}, \boldsymbol{\Psi}_2, g_{\theta_{\mathbf{b}_2}}\right)\\ &= \prod_{i=1}^{m} 2^{n_i} f\left(\mathbf{y}_i|\mathbf{X}'_{1i}\boldsymbol{\beta}_1 + \mathbf{Z}'_{1i}\mathbf{b}_{1i}, \sigma^2_\epsilon\mathbf{I}_{n_i}\right) 2^{q_1} f\left(\mathbf{b}_{1i}|\boldsymbol{\mu}_{\mathbf{b}_1}, \boldsymbol{\Psi}_1, \mathbf{D}_{\mathbf{b}_1}, g_{\theta_{\mathbf{b}_1}}\right)\\ &\quad \cdot \int_0^\infty f\left(\nu_{1i}|\boldsymbol{\mu}_{\nu_{1i}}, \boldsymbol{\Sigma}_{\nu_{1i}}; g_{\theta_{\mathbf{b}_1}}\right) d\nu_{1i} \prod_{i=1}^{n}\left\{h^{\delta_{it}}\left(t_i|\mathbf{X}'_{2i}, \mathbf{Z}'_{2i}, \mathbf{b}_{2i}\right)\right\}\\ &\quad \cdot \exp\left\{-H_0(t_i)\exp\left[\mathbf{X}'_{2i}\boldsymbol{\beta}_2 + \mathbf{Z}'_{2i}\mathbf{b}_{2i}\right]\right\} 2^{q_2} f\left(\mathbf{b}_{2i}|\boldsymbol{\mu}_{\mathbf{b}_2}, \boldsymbol{\Psi}_2, \mathbf{D}_{\mathbf{b}_2}, g_{\theta_{\mathbf{b}_2}}\right)\\ &\quad \cdot \int_0^\infty f\left(\nu_{2i}|\boldsymbol{\mu}_{\nu_{2i}}, \boldsymbol{\Sigma}_{\nu_{2i}}; g_{\theta_{\mathbf{b}_2}}\right) d\nu_{2i}\end{aligned}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2_\epsilon, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \mathbf{D}_{\mathbf{b}_1}, \mathbf{D}_{\mathbf{b}_2}, \nu_{\mathbf{b}_1}, \nu_{\mathbf{b}_2})$ are the unknown parameters.

4. The Bayesian structure

For Bayesian modeling, suitable prior distributions need to be selected for all of the parameters in the model. For the fixed effects model, a multivariate normal distribution is selected. Accordingly, the parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ which are p_1 and p_2 variates are considered to follow the $N_{p_l}(\boldsymbol{\beta}_{\circ l}, \mathbf{S}_{\beta_l})$ for $l=1, 2$. For the scale matrix of random effects $\boldsymbol{\Psi}_l$, the inverse Wishart prior distribution $IW(\boldsymbol{\tau}_{b_l}, \boldsymbol{S}_{b_l})$ is used for $l=1, 2$. The prior distribution of the error-scale parameter σ_ϵ^2 is taken as the inverse gamma distribution with parameters α_ϵ and τ_ϵ. The skewness parameter $\mathbf{d}_b$ is taken to follow the truncated multivariate normal distribution $N_{p_l}(\boldsymbol{\mu}_{d_l}, \boldsymbol{\Sigma}_{d_l})I(\mathbf{d}_b > 0)$ for $l=1, 2$. Furthermore, the uniform or truncated exponential distributions are used as the prior distribution for the degree of freedom parameter of the skew-t distribution.

To obtain parameter estimates, the simultaneous posterior density should be calculated, which is achieved by multiplying the simultaneous likelihood function and simultaneous prior distribution of parameters. By calculating the simultaneous posterior density and obtaining the full conditional posterior densities for individual parameters, we can conclude that these distributions do not have closed form. Accordingly, the direct calculation of the posterior mean of these distributions proves to be difficult and impossible. Hence, Markov Chain Monte Carlo (MCMC) methods are used. According to the observed and prior distributions considered, the following posterior distribution can be obtained

$$
\begin{aligned}
&\pi\left(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{b}_{1i}, \mathbf{b}_{2i}, \mathbf{d}_{\mathbf{b}_1}, \mathbf{d}_{\mathbf{b}_2}, \nu_1, \nu_2, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2 | \mathbf{y}, \mathbf{t}\right) \propto \left\{\prod_{i=1}^{n} \phi\left(\mathbf{y}_i | \mathbf{X}_{1i}' \boldsymbol{\beta}_1 + \mathbf{Z}_{1i}' \mathbf{b}_{1i}, \sigma_\epsilon^2 \mathbf{I}_{n_i}\right)\right. \\
&\cdot \phi_{q_1}\left(\mathbf{b}_{1i} | \boldsymbol{\mu}_{\mathbf{b}_{1i}} + \mathbf{D}_{\mathbf{b}_{1i}}, \mathbf{S}_{\mathbf{b}_{1i}}^{-2} \boldsymbol{\Psi}_1\right) \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{01})' \mathbf{S}_{\beta_1}^{-1} (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{01})\right\} \\
&\cdot \exp\left\{-\frac{1}{2}\left(\mathbf{d}_{\mathbf{b}_{1i}} - \boldsymbol{\mu}_{\mathbf{d}_1}\right)' \boldsymbol{\Sigma}_{\mathbf{d}_1}^{-1}\left(\mathbf{d}_{\mathbf{b}_{1i}} - \boldsymbol{\mu}_{\mathbf{d}_1}\right)\right\} I(\mathbf{d}_{\mathbf{b}_{1i}} > 0)\Big\} \exp\left\{-\frac{\nu_{\mathbf{b}_{1i}}}{\lambda_{\mathbf{b}_{1i}}}\right\} I(\nu_{\mathbf{b}_{1i}} > 2) \\
&\cdot \prod_{i=1}^{n} h^{\delta_{it}}(t_i | \mathbf{X}_{2i}, \mathbf{Z}_{2i}, \mathbf{b}_{2i}) \exp\left\{-H_0(t_i) \exp\left(\mathbf{X}_{2i}' \boldsymbol{\beta}_2 + \mathbf{Z}_{2i}' \mathbf{b}_{2i}\right)\right\} \\
&\cdot \phi_{q_2}\left(\mathbf{b}_{2i} | \boldsymbol{\mu}_{\mathbf{b}_{2i}} + \mathbf{D}_{\mathbf{b}_{2i}} \mathbf{w}_{\mathbf{b}_{2i}}, \mathbf{S}_{\mathbf{b}_{2i}}^{-2} \boldsymbol{\Psi}_2\right) \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_{02})' \mathbf{S}_{\beta_2}^{-1} (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_{02})\right\} \\
&\cdot \exp\left\{-\frac{1}{2}\left(\mathbf{d}_{\mathbf{b}_{2i}} - \boldsymbol{\mu}_{\mathbf{d}_2}\right)' \boldsymbol{\Sigma}_{\mathbf{d}_2}^{-1}\left(\mathbf{d}_{\mathbf{b}_{2i}} - \boldsymbol{\mu}_{\mathbf{d}_2}\right)\right\} I(\mathbf{d}_{\mathbf{b}_{2i}} > 0) \exp\left\{-\frac{\nu_{\mathbf{b}_{2i}}}{\lambda_{\mathbf{b}_{2i}}}\right\} I(\nu_{\mathbf{b}_{2i}} > 2).
\end{aligned}
$$

The use of the MCMC methods is contingent upon obtaining full conditional posterior distributions for the parameters of the model. If we denote by $\boldsymbol{\Theta}_{(-\boldsymbol{\theta})}$ a vector parameters with $\boldsymbol{\theta}$ omitted then

$$
\boldsymbol{\beta}_1 | \boldsymbol{b}_{1i}, \mathbf{d}_{\mathbf{b}_{1i}}, \nu_{\mathbf{b}_{1i}}, \sigma_\epsilon^2, \mathbf{S}_{\mathbf{b}_{1i}}^2, \mathbf{y} \sim N_{p_1}\left(\boldsymbol{\mu}_{\beta_1}^{-1} \alpha_{\beta_1}, \boldsymbol{\mu}_{\beta_1}^{-1}\right)
$$

where

$$
\alpha_{\boldsymbol{\beta}_1} = \mathbf{S}_{\beta_1}^{-1} \boldsymbol{\beta}_{01} + \sum_{i=1}^{n} \sigma_\epsilon^2 \mathbf{X}_{1i}' (\mathbf{y}_i - \mathbf{Z}_{1i}' \mathbf{b}_{1i})
$$

and

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_1}^{-1} = \sum_{i=1}^{n} \frac{1}{\sigma_\epsilon^2} \mathbf{X}_{1i}' \mathbf{I}_{n_i}^{-1} \mathbf{X}_{1i} + \mathbf{S}_{\boldsymbol{\beta}_1}^{-1}$$

Furthermore,

$$\mathbf{b}_{1i} | \boldsymbol{\Theta}_{(-\mathbf{b}_{1i})}, \mathbf{y}, \mathbf{t} \sim N_q \left(\boldsymbol{\Omega}_{\mathbf{b}_{1i}}^{-1} \alpha_{\mathbf{b}_{1i}}, \boldsymbol{\Omega}_{\mathbf{b}_{1i}}^{-1} \right)$$

where

$$\boldsymbol{\Omega}_{\mathbf{b}_{1i}}^{-1} = \mathbf{S}_{\mathbf{b}_{1i}}^2 \boldsymbol{\Psi}_1 + \sigma_\epsilon^{-2} \mathbf{Z}_{1i}' \mathbf{I}_{n_i}^{-1} \mathbf{Z}_{1i}$$

and

$$\alpha_{\mathbf{b}_{1i}} = \sigma_\epsilon^{-2} \mathbf{Z}_{1i}' \mathbf{I}_{n_i}^{-1} \left(\mathbf{y}_i - \mathbf{X}_{1i}' \boldsymbol{\beta}_1 \right) + \Psi_1^{-1} \left(\mathbf{D}_{\mathbf{b}_{1i}} \nu_{\mathbf{b}_{1i}} - \boldsymbol{\mu}_{\mathbf{b}_{1i}} \right)$$

Furthermore,

$$\nu_{\mathbf{b}_{1i}} | \boldsymbol{\Theta}_{(-\nu_{\mathbf{b}_{1i}})} \sim N_{q_1} \left(\boldsymbol{\Omega}_{\nu_{\mathbf{b}_{1i}}}^{-1} \alpha_{\nu_{\mathbf{b}_{1i}}}, \boldsymbol{\Omega}_{\nu_{\mathbf{b}_{1i}}}^{-1} \right) I(\nu_{\mathbf{b}_{1i}} > 0)$$

where

$$\boldsymbol{\Omega}_{\nu_{\mathbf{b}_{1i}}}^{-1} = \mathbf{S}_{\mathbf{b}_{1i}}^2 (\mathbf{D}_{\mathbf{b}_{1i}} \boldsymbol{\Psi}_1 \mathbf{D}_{\mathbf{b}_{1i}}) + \mathbf{I}_{q_1}$$

Furthermore,

$$\boldsymbol{\Psi}_1 | \boldsymbol{\Theta}_{(-\boldsymbol{\Psi}_1)}, \mathbf{y}, \mathbf{t} \sim IW_{q_1} (n + \alpha_{\mathbf{b}_{1i}}, \boldsymbol{\Omega}_{\boldsymbol{\Psi}_1})$$

where

$$\boldsymbol{\Omega}_{\boldsymbol{\Psi}_1} = \mathbf{S}_{\mathbf{b}_{1i}}^{-1} + \sum_{i=1}^{n} \mathbf{S}_{\mathbf{b}_{1i}}^2 \left(\mathbf{b}_{1i} - \mathbf{D}_{\mathbf{b}_{1i}} \nu_{\mathbf{b}_{1i}} + \boldsymbol{\mu}_{\mathbf{b}_{1i}} \right) \left(\mathbf{b}_{1i} - \mathbf{D}_{\mathbf{b}_{1i}} \nu_{\mathbf{b}_{1i}} + \boldsymbol{\mu}_{\mathbf{b}_{1i}} \right)'$$

The conditional posterior distribution for $\boldsymbol{\Psi}_2$ can be obtained similarly. The conditional distributions for other parameters are not in closed-form or well-known.

5. Simulation

This section performs simulation for scrutinizing the performance of the proposed model in the presence and absence of skewness. A sample with the size of n with 10000 repetitions was generated and the mixed effect model was considered as

$$y_{ij} = \beta_{11} + \beta_{12} t_j + \beta_{13} x_i + b_{1i} + b_{2i} t_j + \epsilon_{ij}$$

where $i = 1, 2, \ldots, n$, $j = 1, 2, 3$ and $\mathbf{t} = (3, 6, 9)$. Moreover, ϵ_{ij}'s were generated from the $N(0, 1)$ distribution. The x_i's were generated from the Bernoulli (0.3) distribution. The following Weibull proportional hazards model was considered:

$$h(t_i) = \lambda t_i^{\lambda - 1} \exp \{ \beta_{21} + \beta_{22} x_i + r_1 b_{1i} + r_2 b_{2i} \}$$

Moreover, for the random effects, one of $N(0, 1)$, t distribution, or skew-t distribution was considered. The simulation results are provided in Figures 2 and 3. N – N denotes the model where residuals are $N(0, 1)$ and random effects are also

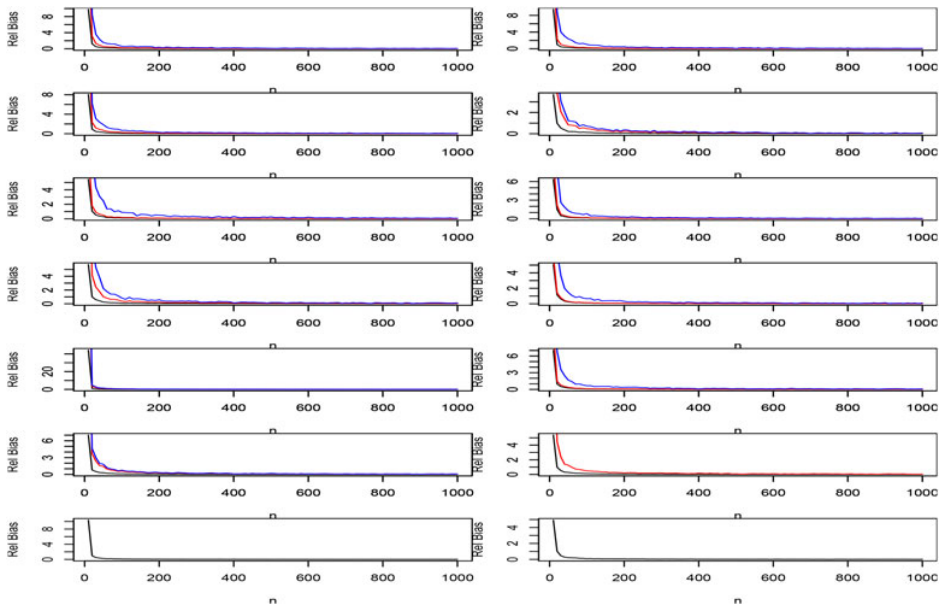


Figure 2. Relative biases versus n of the parameter estimates under the $N-N$ (blue), $N-T$ (red) and $N-ST$ (black) models. The parameters are in the following order: β_{11} (first row, left), β_{12} (first row, right), β_{13} (second row, left), β_{21} (second row, right), β_{22} (third row, left), d_{11} (third row, right), d_{12} (fourth row, left), d_{22} (fourth row, right), λ (fifth row, left), r_1 (fifth row, right), r_2 (sixth row, left), ν (sixth row, right), ω_1 (seventh row, left) and ω_2 (seventh row, right).

$N(0, 1)$. $N-T$ denotes the model where residuals are $N(0, 1)$ and random effects are t distributed with degree of freedom ν. $N-ST$ denotes the model where residuals are $N(0, 1)$ and random effects are skew-t distributed with skewness parameters (ω_1, ω_2) and degree of freedom parameter ν. d_{11}, d_{12}, d_{21} and d_{22} denote elements of the covariance matrix.

Figures 2 and 3 plot the relative biases and relative mean squared errors of the estimators of the parameters for the $N-N$, $N-T$ and $N-ST$ models. Note that the parameters for the $N-N$ model are β_{11}, β_{12}, β_{13}, β_{21}, β_{22}, d_{11}, d_{12}, d_{22}, λ, r_1 and r_2. The parameters for the $N-T$ model are β_{11}, β_{12}, β_{13}, β_{21}, β_{22}, d_{11}, d_{12}, d_{22}, λ, r_1, r_2 and ν. The parameters for the $N-ST$ model are β_{11}, β_{12}, β_{13}, β_{21}, β_{22}, d_{11}, d_{12}, d_{22}, λ, r_1, r_2, ν, ω_1 and ω_2. In the simulations, the parameter values were set as $\beta_{11}=4, \beta_{21}=\beta_{22}=1$, $\beta_{12}=\beta_{13}=1$, $r_1=r_2=-1$, $\lambda=2$, $\omega_1=$ -5, $\omega_2=4$, $\nu=5$, $d_{11}=3, d_{12}=0$ and $d_{22}=4$. The x and y axes in Figure 3 are in logarithmic scale.

The $N-N$, $N-T$ and $N-ST$ models were fitted to each of the 10000 samples of size n. This yields 10000 estimates of the parameters in the $N-N$, $N-T$ and $N-ST$ models. For example, for the $N-N$ model, we obtained the estimates $\hat{\beta}_{11,n,i}, \hat{\beta}_{12,n,i}$, $\hat{\beta}_{13,n,i}, \hat{\beta}_{21,n,i}$, $\hat{\beta}_{22,n,i}, \hat{d}_{11,n,i}$, $\hat{d}_{12,n,i}, \hat{d}_{22,n,i}$, $\hat{\lambda}_{n,i}$, $\hat{r}_{1,n,i}$ and $\hat{r}_{2,n,i}$ for $i=1, 2, \ldots, 10000$ and each n. These estimates were used to compute the relative biases and relative mean squared errors of the estimators for each n and for the three models. For example, for the $N-N$ model, we computed the relative biases and relative mean squared errors as

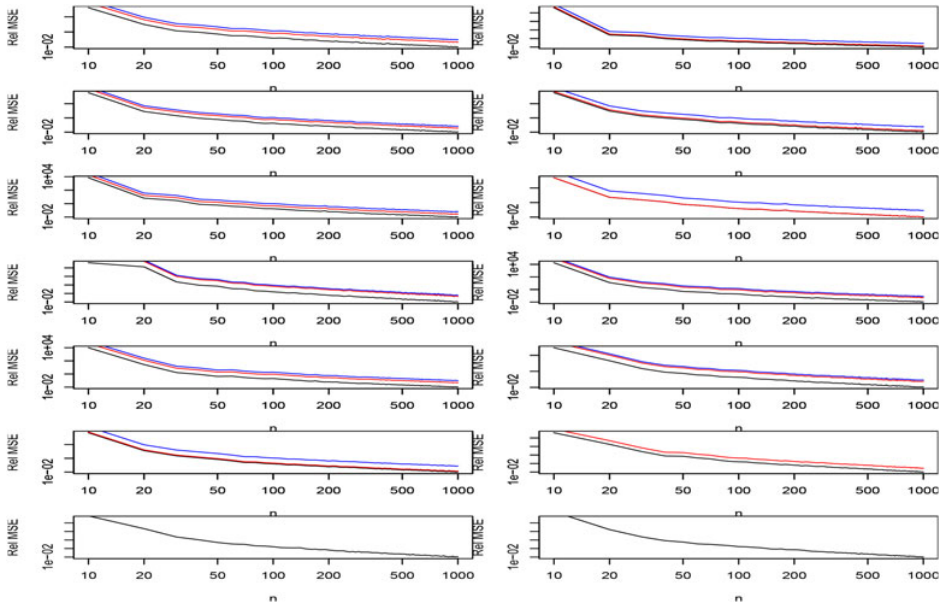


Figure 3. Relative mean squared errors versus n of the parameter estimates under the $N - N$ (blue), $N - T$ (red) and $N - ST$ (black) models. The parameters are in the following order: β_{11} (first row, left), β_{12} (first row, right), β_{13} (second row, left), β_{21} (second row, right), β_{22} (third row, left), d_{11} (third row, right), d_{12} (fourth row, left), d_{22} (fourth row, right), λ (fifth row, left), r_1 (fifth row, right), r_2 (sixth row, left), ν (sixth row, right), ω_1 (seventh row, left) and ω_2 (seventh row, right).

$$\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\beta}_{11,n,i}-\beta_{11}\right)/\beta_{11}, \quad \frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\beta}_{12,n,i}-\beta_{12}\right)/\beta_{12},$$

$$\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\beta}_{13,n,i}-\beta_{13}\right)/\beta_{13}, \quad \frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\beta}_{21,n,i}-\beta_{21}\right)/\beta_{21},$$

$$\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\beta}_{22,n,i}-\beta_{22}\right)/\beta_{22}, \quad \frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{d}_{11,n,i}-d_{11}\right)/d_{11},$$

$$\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{d}_{12,n,i}-d_{12}\right)/d_{12}, \quad \frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{d}_{22,n,i}-d_{22}\right)/d_{22},$$

$$\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\lambda}_{n,i}-\lambda\right)/\lambda, \quad \frac{1}{10000}\sum_{i=1}^{10000}(\hat{r}_{1,n,i}-r_1)/r_1,$$

$$\frac{1}{10000}\sum_{i=1}^{10000}(\hat{r}_{2,n,i}-r_2)/r_2$$

and

$$\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\beta}_{11,n,i}-\beta_{11}\right)^2/\beta_{11}^2, \quad \frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\beta}_{12,n,i}-\beta_{12}\right)^2/\beta_{12}^2,$$
$$\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\beta}_{13,n,i}-\beta_{13}\right)^2/\beta_{13}^2, \quad \frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\beta}_{21,n,i}-\beta_{21}\right)^2/\beta_{21}^2,$$
$$\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\beta}_{22,n,i}-\beta_{22}\right)^2/\beta_{22}^2, \quad \frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{d}_{11,n,i}-d_{11}\right)^2/d_{11}^2,$$
$$\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{d}_{12,n,i}-d_{12}\right)^2/d_{12}^2, \quad \frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{d}_{22,n,i}-d_{22}\right)^2/d_{22}^2,$$
$$\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\lambda}_{n,i}-\lambda\right)^2/\lambda^2, \quad \frac{1}{10000}\sum_{i=1}^{10000}(\hat{r}_{1,n,i}-r_1)^2/r_1^2,$$
$$\frac{1}{10000}\sum_{i=1}^{10000}(\hat{r}_{2,n,i}-r_2)^2/r_2^2$$

respectively, for each n.

The relative biases and the relative mean squared errors generally decrease with n. The relative biases appear generally positive. They appear largest for the $N-N$ model, followed by the $N-T$ model and then the $N-ST$ model. The relative mean squared errors appear largest for the $N-N$ model, followed by the $N-T$ model and then the $N-ST$ model. Hence, the $N-ST$ is the best model and the $N-N$ model is the worst model. For the $N-ST$ model, the relative biases appear reasonably small for all $n \geq 400$ and the relative mean squared errors appear reasonably small for all $n \geq 500$.

In our simulations, we have chosen specific values for $\mathbf{t}$, β_{11}, β_{12}, β_{13}, β_{21}, β_{22}, d_{11}, d_{12}, d_{22}, λ, r_1, r_2, ν, ω_1 and ω_2. But the results of simulation were similar for a wide range of values of these parameters. In particular, the relative biases and the relative mean squared errors generally decreased with n; they always appeared largest for the $N-N$ model; they always appeared smallest for the $N-ST$ model; the relative biases for the $N-ST$ model always appeared reasonably small for all $n \geq 400$; the relative mean squared errors for the $N-ST$ model always appeared reasonably small for all $n \geq 500$.

6. Application

This section analyses data from a longitudinal study conducted on 236 breast-cancer patients. After the surgery, the participants of the study went to a hospital in Tehran. The size of the lymphedema occurring in their arms were recorded at different time intervals after cancer surgery.

First, we tested the data for normality. We used the following tests: Kolmogorov-Smirnov, Shapiro-Wilk's, Jarque-Bera, Anderson-Darling, Cramer-von Mises, Pearson chi-square and Shapiro-Francia tests. All of them returned a p value less than 0.001, giving strong evidence that the data are not normally distributed.

Next, we performed tests for symmetry (or zero skewness) of the data. We used the following tests: Cabilio-Masaro test (Cabilio and Masaro 1996), Mira test (Mira 1999)

Table 1. The criteria for selecting models for the analysis of the breast-cancer data.

Criterion value	Model			
	N-N	N-T	T-T	ST-ST
AIC	10453	8914	7216	6512
BIC	10601	9003	7314	6672

Table 2. The Bayesian estimates of parameters (standard deviation) and 95 percent confidence intervals for the analysis of the breast-cancer data.

Parameter	Estimate	Standard deviation	95% CI
Intercept(β_{11})	10.290	0.725	(8.933,11.791)
Time(β_{12})	–0.306	0.048	(–0.444,-0.244)
Time · meta (β_{13})	0.001	0.071	(–0.118,0.163)
Marital (β_{14})	0.036	0.666	(–1.267,1.269)
Age (β_{15})	–0.326	0.046	(–4.240,-0.242)
Grade (β_{16})	0.186	0.265	(–0.318,0.713)
Intercept (β_{21})	–0.234	0.388	(–0.982,0.542)
Meta (β_{22})	–5.207	0.636	(–6.221,-3.738)
Marital (β_{23})	0.453	0.292	(–0.147,0.983)
Age (β_{24})	0.028	0.657	(–1.279,1.281)
Grad (β_{25})	2.321	0.371	(1.594,3.075)
r_1	–0.313	0.047	(–0.411,-0.225)
r_2	–3.538	0.481	(–4.507,-2.644)
ν_1	4.487	0.461	(7.296,8.987)
ν_2	3.987	0.652	(2.915,5.441)
ω_1	0.452	0.281	(–1.145,0.985)
ω_2	0.861	0.321	(0.161,1.618)

and the MGG test (Miao et al. 2006). All of them returned a p value less than 0.0001, giving strong evidence that the data are skewed.

The following model for the size of the lymphedema is considered:

$$y_{ij} = \beta_{11} + \beta_{12}t_j + \beta_{13}t_i\,\text{meta}_i + \beta_{14}\,\text{marital}_i + \beta_{15}\,\text{age}_i + \beta_{16}\,\text{grad}_i + b_{1i} + b_{2i}t_j + \epsilon_{ij}$$

where the meta variable denotes the Metastasis in the patient, marital denotes the marital state of the patient (single, married, widow, divorced), age refers to the age of the patient, grad denotes tumor's degree of deterioration (low, average, high), and t_{ij} denotes the time of patient visiting the hospital. For the time-to-event occurrence, a Weibull distribution is considered: $T_i|\ \mathbf{b}_i \sim \text{Weibull}(\eta_i^t, \lambda)$ with

$$\eta_i^t = \beta_{21} + \beta_{22}\,\text{meta}_i + \beta_{23}\,\text{marital}_i + \beta_{24}\,\text{age}_i + \beta_{25}\,\text{grad}_i + r_1 b_{1i} + r_2 b_{2i}$$

where $\mathbf{b}_i = (b_{1i}, b_{2i})$ denotes the shared random effect between the two models and assumed to have a skew-t distribution. To run the Bayesian model by considering the distributions introduced in Sections 3 and 4, two parallel Markov Chain Monte Carlo (MCMC) were conducted with different primary values and 100000 repetitions for each run. The first 30000 repetitions were eliminated and posterior inference was based on the remaining 70000 samples. In all of the models $\boldsymbol{\beta}_k \sim N_{p_k}(0, 10000\mathbf{I}_{p_k})$, where $\mathbf{I}_{p-k}$ is a p_k dimensional identity matrix, $r_k \sim N(0, 100), p_1 = 6, p_2 = 5$ and $k = 1, 2$. The prior distribution for the skew-t distribution degree of freedom parameter is $\nu \sim U(1, 9)$. Moreover, the prior distribution for the skewness parameters is $N(2, 1)I(\delta > 0)$. Table 1 compares the Bayesian models, AIC and BIC indexes are given.

According to both AIC and BIC, the ST-ST model is the most appropriate one. Table 2 gives parameter estimates (standard deviation in brackets) and 95 percent confidence intervals for the ST-ST model.

The estimates in Table 2 can be interpreted as follows: a unit change in marital status increases the size of lymphedema by 0.036 and increases the time to event by 0.453; a unit increase in age decreases the size of lymphedema by 0.326 and increases the time to event by 0.028; a unit change in tumor's degree of deterioration increases the size of lymphedema by 0.186 and increases the time to event by 2.321; and so on. The estimates of the following parameters do not appear significantly different from zero: β_{13}, β_{16}, β_{21}, β_{23}, β_{24} and ω_1. The estimates of the following parameters appear significantly negative: β_{12}, β_{15}, β_{22}, r_1 and r_2. The estimates of the following parameters appear significantly positive: β_{11}, β_{25}, ν_1, ν_2 and ω_2.

The prior distributions were selected as $U(1,\ 9)$ and $N(2,1)I(\delta>0)$ for convenience. But other choices gave similar results in that the $ST - ST$ model always showed the best performance.

7. Conclusions

In the previous sections, simultaneous modeling of the longitudinal and survival data for the breast-cancer data was investigated under the assumptions of skewed and symmetric distributions. For the longitudinal data modeling, a linear-mixed model was used. For the survival data modeling, the Weibull proportional hazards model was used. To estimate the parameters, because of the form of the likelihood function, the maximum likelihood method could not be used directly. Furthermore, to increase the exactness of the estimates, the Bayesian method was used. By comparing different distributional assumptions for residuals and random effects, a model with the skew-t (ST) distribution assumption was proved to be the most efficient. This conclusion was also obtained in the simulation part.

Some future are work to apply other multivariate skew models to the data discussed in Section 6. Possible choices may include mixtures of multivariate skew-t distributions due to Lee and McLachlan (2014); the multivariate skew-normal generalized hyperbolic distribution due to Vilca et al. (2014); the multivariate Birnbaum-Saunders distribution due to Jamalizadeh and Kundu (2015); the multivariate skew-normal-Cauchy distribution due to Kahrari et al. (2016); the multivariate geometric skew-normal distribution due to Kundu (2017); scale and shape mixtures of multivariate skew-normal distributions due to Arellano-Valle et al. (2018); the skew mixed effects model due to Eftekhari Mahabadi and Rahimi Jafari (2018); the multivariate skew slash distribution due to Tian et al. (2018).

Acknowledgments

The authors would like to thank the Editor and the two referees for careful reading and comments which greatly improved the paper.

References

Albert, P., and J. Shih. 2010. On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics* 66 (3):983–7. doi:10.1111/j.1541-0420.2009.01324_1.x.

Arellano-Valle, R. B., H. Bolfarine, and V. H. Lachos. 2007. Bayesian inferences for skew-normal linear mixed models. *Journal of Data Science* 13:415–38.

Arellano-Valle, R. B., C. S. Ferreira, and M. G. Genton. 2018. Scale and shape mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis* 166:98–110. doi:10.1016/j.jmva.2018.02.007.

Azzalini, A. 1985. A class of distribution which includes the normal ones. *Scandinavian Journal Statistics* 12:199–208.

Azzalini, A., and A. Capitanio. 1999. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61 (3): 579–602. doi:10.1111/1467-9868.00194.

Azzalini, A., and A. Dalla-Valle. 1996. The multivariate skew-normal distribution. *Biometrika* 83 (4):715–26. doi:10.1093/biomet/83.4.715.

Baghfalaki, T., and M. Ganjali. 2015. A Bayesian approach for joint modeling of skew-normal longitudinal measurements and time to event data. *REVSTAT-Statistical Journal* 13 (2):169–91.

Baghfalaki, T., M. Ganjali, and R. Hashemi. 2014. Bayesian joint modeling of longitudinal measurements and time-to-event data using robust distributions. *Journal of Biopharmaceutical Statistics* 24 (4):834–5. doi:10.1080/10543406.2014.903657.

Bakar, M. R. A., K. A. Salah, N. A. Ibrahim, and K. Haron. 2009. Bayesian approach for joint longitudinal and time-to-event data with survival fraction. *Bulletin of the Malaysian Mathematical Sciences Society* 32:75–100.

Branco, M. D., and D. K. Dey. 2001. A general class of multivariate skew-normal distributions. *Journal of Multivariate Analysis* 79 (1):99–113. doi:10.1006/jmva.2000.1960.

Brown, E. R., and J. G. Ibrahim. 2003. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* 59:221–8.

Cabilio, P., and J. Masaro. 1996. A simple test of symmetry about an unknown median. *Canadian Journal of Statistics* 24 (3):349–61. doi:10.2307/3315744.

Chi, Y. Y., and J. G. Ibrahim. 2007. Bayesian approach to joint longitudinal and survival models accommodating both zero and nonzero cure fractions. *Statistica Sinica* 17:445–62.

Choudhary, P. K., D. Sengupta, and P. Cassey. 2014. A general skew-t mixed model that allows different degrees of freedom for random effects and error distributions. *Journal of Statistical Planning and Inference* 147:235–47. doi:10.1016/j.jspi.2013.11.015.

Eftekhari Mahabadi, S., and E. Rahimi Jafari. 2018. Skew-mixed effects model for multivariate longitudinal data with categorical outcomes and missingness. *Journal of Applied Statistics* 45 (12):2182–201. doi:10.1080/02664763.2017.1413076.

Faucett, C. L., and D. C. Thomas. 1996. Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine* 15 (15): 1663–85. doi:10.1002/(SICI)1097-0258(19960815)15:15<1663::AID-SIM294>3.0.CO;2-1.

Hashemi, R., H. Jacqmin-Gadda, and D. Commenges. 2003. A latent process model for joint modeling of events and marker. *Lifetime Data Analysis* 9 (4):331–43. doi:10.1023/B:LIDA.0000012420.36627.a6.

Henderson, R., P. Diggle, and A. Dobson. 2000. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 1 (4):465–80. doi:10.1093/biostatistics/1.4.465.

Huang, Y., R. Chen, and G. Dagne. 2011. Simultaneous Bayesian inference for linear, nonlinear and semiparametric mixed-effects models with skew-normality and measurement errors in covariates. *The International Journal of Biostatistics* 7 (1):1. Article 8. doi:10.2202/1557-4679.1292.

Ibrahim, J. G., M. H. Chen, and D. Sinha. 2004. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica pp Sinica*. doi:10.5705/ss.202016.0319.

Jamalizadeh, A., and D. Kundu. 2015. A multivariate Birnbaum-Saunders distribution based on the multivariate skew normal distribution. *Journal of the Japan Statistical Society* 45 (1):1–20. doi:10.14490/jjss.45.1.

Kahrari, F., M. Rezaei, F. Yousefzadeh, and R. B. Arellano-Valle. 2016. On the multivariate skew-normal-Cauchy distribution. *Statistics and Probability Letters* 117:80–8. doi:10.1016/j.spl.2016.05.005.

Kundu, D. 2017. Multivariate geometric skew-normal distribution. *Statistics* 51 (6):1377–97. doi:10.1080/02331888.2017.1355369.

Lange, K. L., R. J. A. Little, and J. M. G. Taylor. 1989. Robust statistical modeling using the *t* distribution. *Journal of the American Statistical Association* 84:881–96. doi:10.2307/2290063.

Lee, S., and G. J. McLachlan. 2014. Finite mixtures of multivariate skew t-distributions: Some recent and new results. *Statistics and Computing* 24 (2):181–202. doi:10.1007/s11222-012-9362-4.

Li, N., R. M. Elashoff, and G. Li. 2009. Robust joint modeling of longitudinal measurements and competing risks failure time data. *Biometrical Journal* 51 (1):19–30. doi:10.1002/bimj.200810491.

Lin, I., J. Ho, and L. Chen. 2009. Analysis of multivariate skew-normal models with incomplete data. *Journal of Multivariate Analysis* 100 (10):2337–51. doi:10.1016/j.jmva.2009.07.005.

Lin, T., and J. Lee. 2008. Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in Medicine* 27 (9):1490–507. doi:10.1002/sim.3026.

Lu, T., M. Lu, M. Wang, J. Zhang, G. H. Dong, and Y. Xu. 2017. Partially linear mixed effects joint models for skewed and missing longitudinal competing risks outcomes. *Journal of Biopharmaceutical Statistics* 18:1–19. doi:10.1080/10543406.2017.1378663.

Miao, W., Y. R. Gel, and J. L. Gastwirth. 2006. A new test of symmetry about an unknown median. In *Random walk, sequential analysis and related topics – a Festschrift in honor of Yuan-Shih Chow*, ed. A. Hsiung, C.-H. Zhang, and Z. Ying, 199–214. Singapore: World Scientific Publisher.

Mira, A. 1999. Distribution-free test for symmetry based on Bonferroni's measure. *Journal of Applied Statistics* 26 (8):959–72. doi:10.1080/02664769921963.

Rizopoulos, D. 2013. *Joint modeling of longitudinal and time-to-event data: Challenges and future directions*. New York: Springer Verlag.

Rizopoulos, D. 2010. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software* 35 (9):1–33. doi:10.18637/jss.v035.i09.

Rizopoulos, D., and P. Ghosh. 2011. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine* 30 (12):1366–80. doi:10.1002/sim.4205.

Rizopoulos, D., G. Verbeke, E. Lesaffre, and Y. Vanrenterghem. 2008a. A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics* 64 (2):611–9. doi:10.1111/j.1541-0420.2007.00894.x.

Rizopoulos, D., G. Verbeke, and G. Molenberghs. 2008b. Shared parameter models under random effects misspecification. *Biometrika* 95 (1):63–74. doi:10.1093/biomet/asm087.

Sahu, S. K., D. K. Dey, and M. Branco. 2003. A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics* 31 (2):129–50. doi:10.2307/3316064.

Song, X., M. Davidian, and A. A. Tsiatis. 2002. A semiparametric likelihood approach to joint modeling longitudinal and time-to-event data. *Biometrics* 58 (4):742–53. doi:10.1111/j.0006-341X.2002.00742.x.

Tian, W., T. Wang, and A. K. Gupta. 2018. A new family of multivariate skew slash distribution. *Communications in Statistics: Theory Methods* 47 (23):5812–24. doi:10.1080/03610926.2017.1402049.

Tsiatis, A. A., and M. Dividian. 2001. A semiparametric estimator for the proportional hazards model with longitudinal covariates measures with error. *Biometrika* 88 (2):447–58. doi:10.1093/biomet/88.2.447.

Tsiatis, A. A., and M. Davidian. 2004. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* 14:809–34.

Tsiatis, A. A., V. Degruttola, and M. S. Wulfsohn. 1995. Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* 90 (429):27–37. doi:10.1080/01621459.1995.10476485.

Vilca, F., N. Balakrishnan, and C. B. Zeller. 2014. Multivariate skew-normal generalized hyperbolic distribution and its properties. *Journal of Multivariate Analysis* 128:73–85. doi:10.1016/j.jmva.2014.03.002.

Wu, L., and R. J. Cook. 2012. Misspecification of Cox regression models with composite endpoints. *Statistics in Medicine* 31 (28):3545. doi:10.1002/sim.5436.

Wulfsohn, M. S., and A. A. Tsiatis. 1997. A joint model for survival and longitudinal data measured with error. *Biometrics* 53 (1):330–9.

Ye, W., X. Lin, and J. M. Taylor. 2008a. A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. *Statistics and Its Interface* 1:33–45. doi:10.4310/SII.2008.v1.n1.a4.

Ye, W., X. Lin, and J. M. Taylor. 2008. Semiparametric modeling of longitudinal measurements and time-to-event data-a two-stage regression calibration approach. *Biometrics* 64 (4):1238–46. doi:10.1111/j.1541-0420.2007.00983.x.